Dissecting tail-anchored protein localization signals with iPAL (imaging pooled-to-array library) scanning

Dissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

am Fachbereich Biologie

der Johannes Gutenberg-Universität Mainz

vorgelegt von

Elena Ivanova

geboren am 27.06.1994

in Sankt Petersburg, Russland

Mainz, August 2022

Dekan:

1. Gutachter:

2. Gutachter:

Tag der mündlichen Prüfung: 18.10.2022

Preface

1 TABLE OF CONTENT

2 SUMMARY

Most proteins in a eukaryotic cell are synthesized by ribosomes in the cytosol and then need to be selectively targeted to different subcellular compartments to fulfill their functions. The core of protein targeting is the recognition of a targeting signal within the nascent protein by a targeting factor for the destination organelle. It is a complex process that involves multiple overlapping pathways and errors in protein targeting occur even under optimal conditions, leading to protein mislocalization. Mutations in localization signals can also result in protein mislocalization and cause severe genetic disorders. Thus, it is important to understand the nature of localization signals, the mechanisms of protein targeting, and the systems involved in the quality control (QC) of protein targeting.

In this study, I describe iPAL scanning (imaging pooled-to-arrayed libraries), a deep mutational scanning approach to dissect protein localization signals. In iPAL, pooled libraries of protein variants with alterations in a potential localization signal are constructed by homologous recombination. Each protein variant is fused to a fluorescent protein (mNeonGreen) to read out its localization with fluorescence microscopy. Using *in vivo* barcoding and deep sequencing, pooled libraries are converted into verified arrayed libraries, followed by high-throughput fluorescence microscopy to determine the localization of each variant. Importantly, iPAL is not limited by the number of phenotypic bins and thus can be used to dissect complex localization signals. Due to the arrayed library format, all generated protein variants can be directly used for downstream applications, e.g., co-localization experiments or genetic screens.

I applied iPAL scanning to investigate localization determinants of tail-anchored (TA) proteins in budding yeast. TA proteins carry a single α-helical transmembrane domain (TMD) at their C-terminus, which enables them to localize to the endoplasmic reticulum (ER), mitochondria, Golgi, nuclear envelope, vacuole and plasma membrane (PM). However, the features of these TMDs crucial for targeting specificity are not well defined. Using iPAL, I generated 1350 unique TMDs with variations in amino acid composition, TMD length and charge of the flanking residues. By analyzing the localization and biophysical properties of TMD variants, I found that TMDs of most ER-resident TA proteins of average hydrophobicity are alone sufficient for correct localization. In contrast, a combination of properties ensures mitochondrial and PM localization. Mitochondrial localization requires low hydrophobicity of the TMD and a high positive charge at the C-terminal extension. Two different types of features lead to PM localization: long TMDs of very high hydrophobicity flanked with positively charged residues at the N-terminus or short cysteine-rich transmembrane modules (CYSTM) of low hydrophobicity. Long length and high hydrophobicity of the TMDs from the first group might be related to the greater thickness of the PM and positively charge flanking residues might be required for the electrostatic interaction with phosphatidylinositides at the PM. It is possible that cysteine residues are important for the oligomerization of CYSTM proteins or for their interaction with TMDs from other PM proteins in the bilayer. However, it remains a

possibility that CYSTM proteins are not integrated into the PM but are associated only peripherally.

In a second line of research, I combined iPAL with genetic screens to demonstrate the downstream application of iPAL and investigate which TMD features are recognized by protein targeting and QC systems. I performed a screen for protein turnover factors by introducing the knockouts of Get3 and Emc3 targeting factors Tul1, Hrd1, Asi1 and Doa10 E3 ligases and Msp1 and Spf1 protein dislocases into an mNeonGreen-TMD library. I observed that substrates of ER and mitochondrial protein QC, namely of Doa10, Msp1 and Spf1, tend to have TMDs of low hydrophobicity. TMDs of low hydrophobicity are generally not recognized by the SRP (signal recognition particle) and GET (guided entry of TA protein) targeting factors and, thus, might be more frequently subjected to QC.

Altogether, my work reveals that different combinations of TMD properties, including hydrophobicity, length and presence of positively charged flanking residues, provide the targeting specificity for ER, mitochondrial and plasma membranes. I anticipate that the iPAL approach will have a wide range of applications and help to understand the interplay between protein sequence and visual phenotypes.

3 ZUSAMMENFASSUNG

Die meisten Proteine in einer eukaryontischen Zelle werden von Ribosomen im Zytosol synthetisiert und müssen dann selektiv in verschiedene subzelluläre Kompartimente gelenkt werden, um ihre Funktionen zu erfüllen. Die Grundlage der Protein Targeting ist die Erkennung eines Lokalisierungssignal im naszierenden Protein durch einen Targeting-Faktor für die Zielorganelle. Es handelt sich um einen komplexen Prozess, an dem mehrere sich überschneidende Wege beteiligt sind, und selbst unter optimalen Bedingungen können Fehler bei der Proteinlokalisierung auftreten, die zu einer Fehllokalisierung von Proteinen führen. Mutationen in Lokalisierungssignalen können ebenfalls zu einer Fehllokalisierung von Proteinen führen und schwere genetische Störungen verursachen. Daher ist es wichtig, die Natur der Lokalisierungssignale, die Mechanismen der Proteinlokalisierung und die Systeme zu verstehen, die an der Qualitätskontrolle der Proteinlokalisierung beteiligt sind.

In dieser Studie beschreibe ich das iPAL-Scanning (Imaging Pooled-to-Arrayed Libraries), einen hochauflösenden Mutationsscan-Ansatz zur Untersuchung von Proteinlokalisierungssignalen. Bei iPAL werden gepoolte Bibliotheken von Proteinvarianten mit Veränderungen in einem potenziellen Lokalisierungssignal durch homologe Rekombination erstellt. Jede Proteinvariante wird mit einem fluoreszierenden Protein (mNeonGreen) fusioniert, um ihre Lokalisierung mit Fluoreszenzmikroskopie auszulesen. Mittels In-vivo-Strichkodierung und hochauflösender Sequenzierung werden gepoolte Bibliotheken in verifizierte Array-Bibliotheken umgewandelt, gefolgt von Hochdurchsatz-Fluoreszenzmikroskopie zur Bestimmung der Lokalisierung jeder Variante. Wichtig ist, dass iPAL nicht durch die Anzahl der phänotypischen Bins begrenzt ist und daher zur Aufschlüsselung komplexer Lokalisierungssignale verwendet werden kann. Aufgrund des Formats der Array-Bibliothek können alle erzeugten Proteinvarianten direkt für nachgeschaltete Anwendungen verwendet werden, z. B. für Ko-Lokalisierungsexperimente oder genetische Screens.

Ich habe iPAL-Scanning angewandt, um die Lokalisierungsdeterminanten von schwanzverankerten (TA) Proteinen in knospenden Hefen zu untersuchen. TA-Proteine tragen an ihrem C-Terminus eine einzelne α-helikale Transmembrandomäne (TMD), die es ihnen ermöglicht, sich im endoplasmatischen Retikulum (ER), in den Mitochondrien, im Golgi, in der Kernhülle, in der Vakuole und in der Plasmamembran (PM) zu lokalisieren. Die Merkmale dieser TMDs, die für die Spezifität der Zielstruktur entscheidend sind, sind jedoch nicht gut definiert. Mit iPAL habe ich 1350 einzigartige TMDs mit Variationen in der Aminosäurezusammensetzung, der TMD-Länge und der Ladung der flankierenden Reste erzeugt. Durch die Analyse der Lokalisierung und der biophysikalischen Eigenschaften der TMD-Varianten fand ich heraus, dass TMDs der meisten ER-residenten TA-Proteine mit durchschnittlicher Hydrophobizität allein für eine korrekte Lokalisierung ausreichen. Im Gegensatz dazu gewährleistet eine Kombination von Eigenschaften die mitochondriale und PM-Lokalisierung. Die mitochondriale Lokalisierung erfordert eine geringe Hydrophobizität der TMD und eine hohe positive Ladung an der C-terminalen Verlängerung. Zwei verschiedene

Arten von Merkmalen führen zur PM-Lokalisierung: lange TMDs mit sehr hoher Hydrophobizität, die von positiv geladenen Resten am N-Terminus flankiert werden, oder kurze cysteinreiche Transmembranmodule (CYSTM) mit geringer Hydrophobizität. Die große Länge und hohe Hydrophobizität der TMDs der ersten Gruppe könnte mit der größeren Dicke der PM zusammenhängen, und die positiv geladene flankierende Reste könnten für die elektrostatische Wechselwirkung mit Phosphatidylinositiden an der PM erforderlich sein. Es ist möglich, dass Cysteinreste für die Oligomerisierung von CYSTM-Proteinen oder für ihre Interaktion mit TMDs von anderen PM-Proteinen in der Doppelschicht wichtig sind. Es besteht jedoch auch die Möglichkeit, dass CYSTM-Proteine nicht in die PM integriert sind, sondern periphär assoziiert anliegen.

In einem zweiten Forschungsstrang habe ich iPAL mit genetischen Screens kombiniert, um die nachgeschaltete Anwendung von iPAL zu demonstrieren und zu untersuchen, welche TMD-Merkmale von Protein-Zielbestimmung und Qualitätskontrollsystemen erkannt werden. Ich führte ein Screening nach Faktoren für den Proteinumsatz durch, indem ich Knockouts der Zielbestimmung-Faktoren Get3 und Emc3, der E3-Ligasen Tul1, Hrd1, Asi1 und Doa10 sowie der Proteindislokasen Msp1 und Spf1 in die mNeonGreen-TMD-Bibliothek einführte. Ich habe festgestellt, dass die Substrate der ER- und mitochondrialen Protein-Qualitätskontrolle, d.h. von Doa10, Msp1 und Spf1, in der Regel TMDs mit geringer Hydrophobizität aufweisen. TMDs mit geringer Hydrophobizität werden im Allgemeinen nicht von den SRP (signal recognition particle)- und GET (guided entry of TA protein)-Zielbestimmung-Faktoren erkannt und könnten daher eher der Qualitätskontrolle unterworfen werden.

Zusammengenommen, zeigt meine Arbeit, dass verschiedene Kombinationen von TMD Eigenschaften, inklusive der Hydrophobizität, der Länge und dem Vorhandensein positiv geladener flankierender Reste die Zielbestimmungsspezifität für Membranen des ER, der Mitochondrien und der PM ausmachen. Ich antizipiere, dass der iPAL-Ansatz ein breites Anwendungsspektrum haben und dazu beitragen wird, das Zusammenspiel zwischen Proteinsequenz und visuellen Phänotypen zu verstehen.

4 INTRODUCTION

4.1 PROTEIN TARGETING IN THE CELL

In eukaryotic cells about two-thirds of nascent proteins are selectively targeted from the cytosol to different subcellular compartments, where they are typically translocated across or embedded into a membrane to fulfill their functions (Hegde and Zavodszky, 2019).

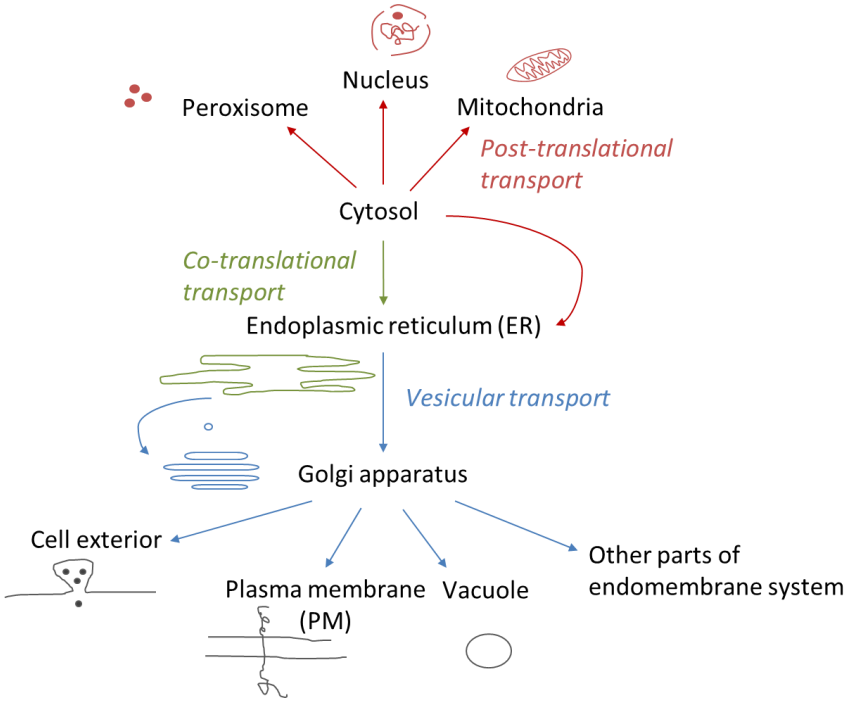

Figure 1. Protein targeting pathways in the cell: co-translational transport, post-translational and vesicular trafficking. Proteins with the N-terminal and internal localization signals are targeted to the ER during translation. Once the localization signal emerges from the ribosome it is recognized by the targeting factor and the ribosome with the nascent chain is brought to ER where the translation is coupled with translocation. Proteins with the C-terminal targeting signals are targeted to the ER post-translationally by the interactions with specific chaperons. A fraction of proteins destined for residence at the PM, endosomes or lysosomes migrate from the ER to these organelles by a vesicular trafficking route. Proteins destined to the nucleus, peroxisome, and mitochondria are targeted to these destinations post-translationally.

Although there are different ways of targeting proteins to their final destination, including co-translational transport, post-translational transport and vesicular trafficking (Figure 1) (Guna et al., 2018; Hegde and Keenan, 2011; Shurtleff et al., 2018), the core of each protein targeting mechanism is the recognition of a targeting signal within the protein by a targeting factor for the destination organelle. The targeting signals for different subcellular compartments may share common features and hence, the recognition by the targeting protein can be imperfect leading to mislocalization even under optimal physiological conditions. Mutations in localization signals can also result in protein mislocalization and cause severe genetic disorders. For instance, mutations in the signal peptides, which lead to the reduced presence of a protein in the correct compartment and the loss of their functions, have been shown to cause familial hypocalciuric hypercalcemia (FHH), neonatal severe hyperparathyroidism (NSHPT) and dentine dysplasia type II (Pidasheva et al., 2005; Rajpar et al., 2002). On the other hand, mutation of the preproinsulin signal sequence has been shown to cause toxicity due to spontaneous gain-of-function of the preproinsulin in the wrong place

promoting β cell death and diabetes (Guo et al., 2014). Thus, it is important to understand the nature of localization signals, the mechanisms of protein targeting, and the systems involved in the quality control of protein targeting. A better understanding of localization signals would help in determining the functional impact of disease-causing mutations and potentially open new therapeutic strategies.

In this section, I will first describe the targeting signals and pathways for transport into different organelles (4.1.1), followed by the features of transmembrane domains important for correct targeting, which is a focus of this study (4.1.2). Co-translational (4.1.3) and post-translational (4.1.4) pathways targeting transmembrane proteins to the ER will be specifically discussed in detail. Some transmembrane proteins destined for residence at the plasma membrane (PM), endosomes or lysosomes migrate from the ER to these organelles by a vesicular trafficking route. The molecular mechanisms of vesicular trafficking have been extensively reviewed elsewhere and will not be described in this work (Bonifacino and Glick, 2004; Liu, 2016; Marshansky and Futai, 2008; Symons and Rusk, 2003; Yorimitsu et al., 2014).

4.1.1 PROTEIN TARGETING SIGNALS

Proteins can contain targeting information within their polypeptide sequence that directs them to the different subcellular compartments. Overall, these localization signals are incompletely understood and diverse but may contain similar features such as positively charged and hydrophobic amino acids (Figure 2). In this chapter, I describe the properties of some types of localization signals and the associated targeting pathways.

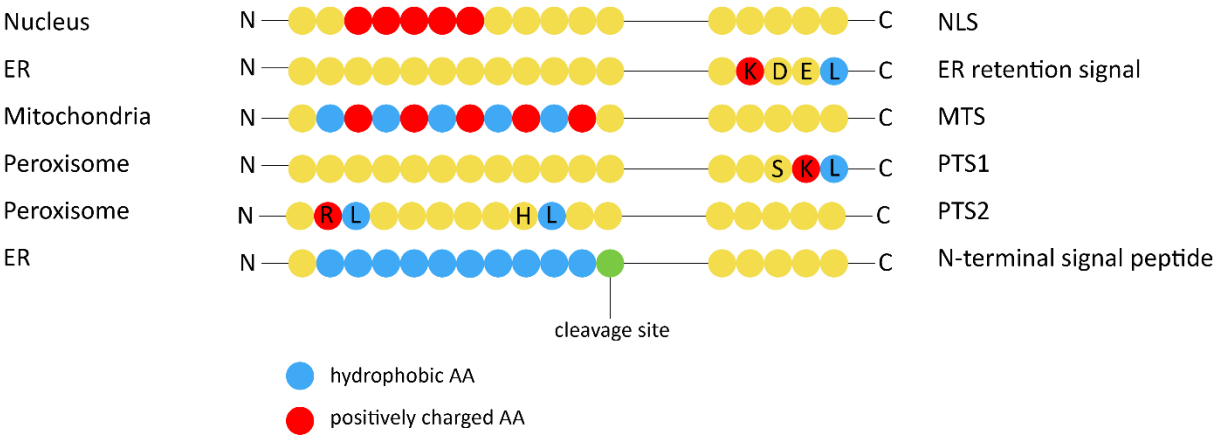

Figure 2. Examples of the different positions of the localization signals within a protein sequence. C, carboxy terminus; N, amino terminus; AA – amino acid. The destinations are listed on the left and targeting signal are listed on the right. Hydrophobic amino acids are indicated in blue and positively charged amino acids in red. NLS – nuclear localization signal; MTS - mitochondrial targeting sequence; PTS - peroxisomal targeting signal.

Nuclear localization signals (NLSs) are one of the best studied targeting signals. NLSs are short peptides that usually consist of lysine and arginine repeats (K/R motif), and target proteins to the nucleus (Kalderon et al., 1984). An NLS is recognized by an importin-α subunit, which in turn is recognized by an importin-β subunit. This complex of the cargo, importin-α and importin-β is then imported to the nucleus after a series of enzymatic steps (Nguyen Ba et al., 2009). Often NLSs are located at protein N-termini. The correct function of proteins with nuclear localization such as transcription factors, DNA repair proteins or chromatin-associated

proteins depends on their NLSs. In the last decades, large-scale experimental datasets were used to develop different prediction tools for NLS identification (Brameier et al., 2007; Lin and Hu, 2013; Nguyen Ba et al., 2009; Requião et al., 2017). This helped to understand that NLSs can be located not only at the N-terminus of the protein (Requião et al., 2017). One example would be nuclear membrane protein Prm3, which has an NLS rather in between C- and N-terminus. Using a deep mutational scanning (DMS) approach Hasle and colleagues performed an analysis of hundreds of NLS variants, which led to an improved NLS predictor and an NLS variant with greatly enhanced nuclear localization (Hasle et al., 2020). This enhanced NLS variant might find broad application in synthetic biology (Hasle et al., 2020). The counterpart of the NLS would be the nuclear export signal, which targets proteins out of the nucleus.

Mitochondrial proteins are mostly synthesized at cytosolic ribosomes and become functional after their insertion into the outer mitochondrial membrane or translocation into the mitochondrion. After translocation via the translocase of the outer membrane (TOM) complex, the proteins can either remain in the intermembrane space or be targeted to the inner mitochondrial membrane or the mitochondrial matrix by translocase of the inner membrane (TIM) complex (Neupert and Herrmann, 2007). A well-studied mitochondria-targeting sequence (MTS) is the matrix targeting sequence of approximately 10-60 amino acids destined for the matrix or the inner mitochondrial membrane. The sequence consists of an alternating pattern of hydrophobic and positively charged amino acids and is located at the N-terminus of the protein. The crucial role of positively-charged residues within the MTS for mitochondrial targeting was discovered using site-directed mutagenesis (Kumamoto et al., 1987). MTS has the potential to form a positively charged amphiphilic α-helix with the segregation of positively charged and hydrophobic residues on opposite faces of the helix. Based on these properties a spectrum of various bioinformatics and experimental tools is available for MTS prediction and analysis (Habib et al., 2007). The MTS is recognized in the cytoplasm by the mitochondrial import stimulation factor (MSF), a cytoplasmic molecular chaperone, which transfers the protein to the mitochondrial translocation machinery. After the MTS-containing protein has been translocated the mitochondrial processing peptidase (MPP) in the mitochondrial matrix may cleave the MTS, which results in a mature mitochondrial protein (Omura, 1998). It has been shown that an ER surface retrieval pathway (ER-SURF) safeguards the import of mitochondrial inner membrane proteins with N-terminal MTS (Hansen et al., 2018). This ER-SURF pathway retrieves mitochondrial proteins from the ER surface and reroutes them to mitochondria with the help of the ER-localized chaperone Djp1, which probably offers an alternative import route for mitochondrial proteins when direct import fails. An additional study demonstrated that the targeting of non-imported mitochondrial carrier proteins is dependent on the guided entry of tail-anchored protein (GET) pathway (Xiao et al., 2021). Upon disruption of GET pathway non-imported mitochondrial proteins are sequestered into protein foci and prevented from the re-targeting from the ER to mitochondria via the ER-SURF pathway.

Interestingly, structural and proteomic data analysis and computational predictions showed that 25% of mitochondrial ribosomal proteins (MRPs), a big class of mitochondrial

proteins, lack an MTS at the N-terminus. A recent experimental study revealed the presence of unconventional inner MTSs for some of the MRPs (Bykov et al., 2022). Similar to a conventional MTS, this inner signal consists of alternating hydrophobic and positively charged amino acids but is localized closer to the core of the protein and cannot be cleaved after the translocation. This inner MTS mediates the binding of MRP to the TOM complex (Bykov et al., 2022). This observation shows that protein targeting is more versatile than expected and further studies are required to complete the map of protein targeting routes in the cell.

Two peroxisomal targeting signals (PTS), namely PTS1 and PTS2, are responsible for directing proteins to the peroxisome. PTS1 consists of three amino acids located at the C-terminus of the protein and can be recognized by the Pex5 receptor (Brocard and Hartig, 2006; El Magraoui et al., 2019; Neuberger et al., 2003). Pex5 can trap PTS1 through the remodeling of its tetratricopeptide repeat (TPR)-domain. Initially, the tripeptide Serine–Lysine–Leucine (SKL) at the extreme C-terminus was identified by site-directed mutagenesis. Further mutagenesis studies resulted in the extended definition of the PTS1-consensus sequence: amino acids at the first position contain a small uncharged side chain (S/A/C), at the penultimate position a positively charged residue (K/R/H) and at the extreme C-terminal position a leucine (L) (Brocard and Hartig, 2006). Further studies on the interaction between Pex5 and PTS1 as well as sequence profile analysis revealed the importance of some amino acids upstream from the tripeptide. Now PTS1 is defined as a dodecamer sequence at the C-terminal end of peroxisomal proteins important for the binding with Pex5 (Brocard and Hartig, 2006).

PTS2 consists of a nine amino acid sequence often present at the N-terminus of the protein and is recognized by the Pex7 receptor (Lazarow, 2006). Initially, a conserved nonapeptide, RLxxxxx(H/Q)L (x=any amino acids), was identified in thiolases and shown to be necessary for the import of these enzymes into peroxisomes (De Hoop and Ab, 1992). Further studies led to the broadening of the PTS2 consensus to (R/K)-(L/V/I)-xxxxx-(H/Q)-(L/A) with the most common PTS2 consensus being R-(L/V/I/Q)-xx-(L/V/I/H)-(L/S/G/A)-x-(H/Q)-(L/A) (Waterham et al., 1994). Pex7, a soluble hydrophilic protein without transmembrane domains, binds the PTS2 nonapeptide and targets proteins to peroxisomes. Cells lacking Pex7 show impaired import of PTS2-containing proteins into peroxisomes (Lazarow, 2006).

The majority of membrane or secretory proteins that are targeted to the ER have N-terminal hydrophobic targeting signals recognized by the signal recognition particle (SRP) during translation (Jomaa et al., 2022). These signals share no obvious amino acid sequence homology but have common biophysical properties. Signal sequences may have different lengths with a hydrophobic core of approximately 8-12 residues flanked by short stretches of polar residues (von Heijne and Abrahmsèn, 1989; Nyathi et al., 2013). They are cleaved by the signal peptidase complex (SPC) during or after protein insertion into the membrane (Paetzel et al., 2002). The translocation pathways targeting these proteins to the ER are described in detail in section 4.1.3. Another type of targeting signal important for ER localization is the four amino acids carboxy-terminal signal HDEL (KDEL in mammals). The presence of KDEL sequence maintains the ER residence of soluble proteins by dynamic retrieval from

downstream compartments of the secretory pathway by the retrograde transport mechanisms. This process is mediated by the p26 KDEL receptor and by the proteins forming the COPI coatomer structure (Stornaiuolo et al., 2003).

4.1.2 TRANSMEMBRANE DOMAINS

Proteins that are permanently attached to the membrane are called integral membrane proteins. Integral membrane proteins with TMDs that span across the lipid bilayer are called transmembrane proteins. TMDs represent a separate class of localization signals. TMDs are not cleavable and provide stable anchoring of the protein in the lipid bilayer as they mostly consist of 16-30 hydrophobic amino acids forming an α-helical structure. The TMD properties ensuring the targeting specificity are the focus of this study and in this chapter (4.1.2) I would like to describe the latest findings about this group of localization signals.

Different types of transmembrane proteins exist (Figure 3). Transmembrane proteins that use their first TMD as both a signal sequence for membrane targeting and a stop-transfer sequence for termination of translocation are called signal-anchored proteins. Their N-terminal TMDs can be recognized during protein synthesis by the SRP and inserted into the ER co-translationally. Type I signal-anchored proteins have cleavable N-terminal signal peptide in addition to the TMD. Type II signal-anchored proteins have TMD located in the middle of the protein sequence, while type III proteins have N-terminal TMD. TMD proteins with small or short cytoplasmic tails at both termini are called correspondingly, small or short TMD proteins. Multipass TMD proteins can have several TMD domains. Transmembrane proteins with a signal-anchored domain at the extreme C- terminus are called C-terminally anchored proteins or TA proteins (Ott and Lingappa, 2002).

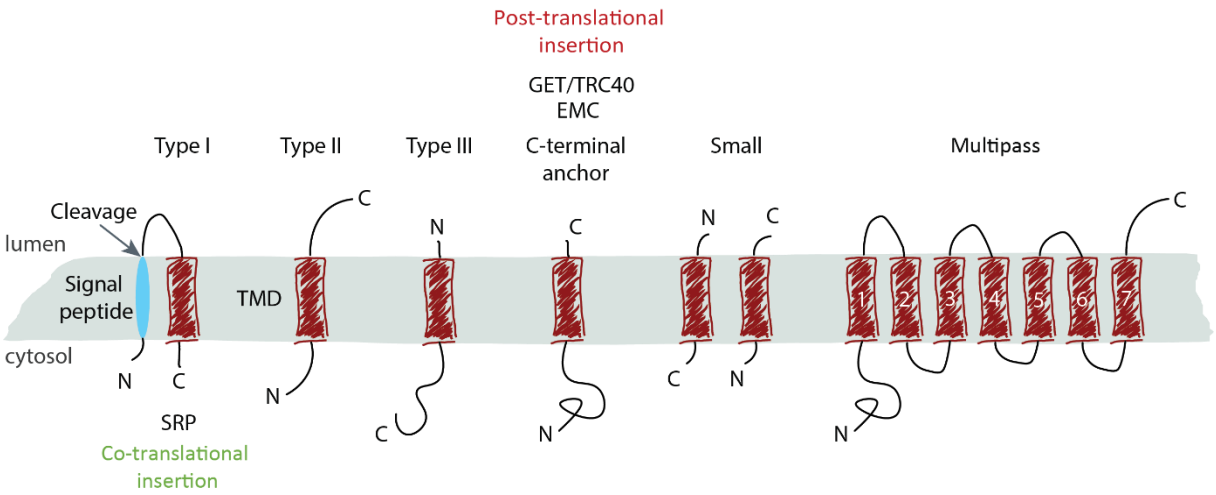

Figure 3. The major classes of transmembrane proteins according to their topology. Type I proteins with the cleavable N-terminal signal peptides; Type II proteins with the TMDs located in the middle of the protein sequence; Type III proteins with the N-terminal TMDs; TA proteins with the TMDs located at the very C-terminus; Small or short transmembrane proteins with very short cytosolic ends; multipass membrane proteins with several TMDs. C, carboxy terminus; N, amino terminus.

TA proteins play roles in various cellular processes, including membrane fusion, protein translocation and apoptosis regulation (Antonsson, 2001; Beilharz et al., 2003). Two

main destinations of TA proteins are ER and mitochondria, although some of them anchor to the Golgi apparatus, nuclear envelope (NE), peroxisome, PM or vacuole (Burri and Lithgow, 2004). C-terminal TMDs of TA proteins are targeted to ER by the GET pathway in yeast and the TRC40/GET pathway in mammals (Denic, 2012; Schuldiner et al., 2008). In mammalian cells a second pathway to insert TA proteins into the ER exists, which is the ER-membrane protein complex (EMC)-dependent pathway (Guna et al., 2018; Hegde and Keenan, 2011). The EMC-dependent pathway also exists in yeast but only a few TA protein substrates were identified (Bai et al., 2020; Güngör et al., 2022; Miller-Vedam et al., 2020). Some TA proteins destined for residence at the plasma membrane (PM), endosomes or lysosomes migrate from the ER to these organelles by a vesicular trafficking route. The mechanisms of TA protein targeting to the mitochondria remain unclear. Details about the ER targeting machinery for signal-anchored and TA proteins can be found in sections 4.1.3 and 4.1.4 respectively.

High hydrophobicity of the C-terminal TMD is required for proper membrane insertion of ER-destined TA proteins (Wang et al., 2010). TMDs of mitochondrial TA proteins tend to have low hydrophobicity and low helical content and are usually flanked by positively-charged arginine and lysine residues (Costello et al., 2017). It has been shown for the mitochondrial TA protein Fis1 that positively charged residues C-terminal to the TMD are important for mitochondrial localization, as in the absence of these residues protein tends to mislocalize to the ER (Habib et al., 2003; Keskin et al., 2017; Rao et al., 2016). Adding positively charged residues to the C-terminal side of the TMD of Bos1, an ER-resident TA protein, slowed its insertion into ER, suggesting in this case that the GET pathway can differentiate and disfavor TMDs with positively-charged C-terminal flanks (Rao et al., 2016). In mammalian cells, adding a very high positive charge to the C-terminal part of the mitochondrial TMD with moderate hydrophobicity directed the protein to peroxisomes, and removing the C-terminal positive charge directed the protein to the ER (Costello et al., 2017). These results suggest an interplay between tail charge and TMD hydrophobicity in organelle targeting. However, not all mitochondrial TA proteins possess a positive C-terminal charge and it is not clear if there are any additional features required for the correct targeting to mitochondria.

Computational analysis of the TMDs from all single-span transmembrane proteins revealed that TMDs have organelle-specific properties, most likely related to the differences in membrane composition (Figure 4) (Sharpe et al., 2010). Thus, plasma membrane (PM) proteins have longer TMDs, which correlates with the greater thickness of the PM compared to the other compartments (Figure 4). It has been shown that the thickness of a membrane depends on the presence of lipids such as sterols or sphingolipids, which increase acyl chain order and thus thicken the bilayer (Brown and London, 1998; Lewis and Engelman, 1983). Compared to the ER and Golgi, the PM is enriched in sterols and sphingolipids and thus, is expected to have a greater thickness (Holthuis et al., 2001). Sharpe and colleagues demonstrated that the TMDs of PM proteins have an asymmetry in the distribution of valine, glycine, and leucine, with valine and glycine being favored in more exoplasmic positions, whereas leucine shows the asymmetry towards the cytosolic end (Figure 4). This asymmetry in the amino acid composition is linked to the residue volume. Several studies of individual

proteins demonstrated that the alterations of the TMD sequence lead to changes in localization. For instance, the elongation of the TMD of the ER-resident TA protein Ubc6 from 17 to 21 amino acids led to a change in localization from ER to Golgi, and further elongation from 17 to 26 amino acids allowed the modified protein to be localized to the PM and vacuoles (Yang et al., 1997). In another study, swapping the TMD of the ER-resident TA protein Ufe1 with the longer TMD from the PM resident TA protein Sso1 resulted in the localization of the chimeric protein at the PM (Rayner and Pelham, 1997). By creating chimeric TMDs with different exoplasmic residue volumes, Quiroga and colleagues observed that short TMDs with high-volume exoplasmic hemi-TMDs confer Golgi membrane residence, whereas TMDs with low-volume exoplasmic hemi-TMDs, either short or long, confer plasma membrane residence to these proteins (Quiroga et al., 2013). Increasing the residue volume of the exoplasmic hemi-TMD from the PM TA protein Sso1 leads to the polarized distribution of the chimeric protein at the PM. TMD-dependent sorting in the endocytic pathway was also observed for the vesicle membrane receptor protein (v-SNARE) Snc1 (Lewis et al., 2000).

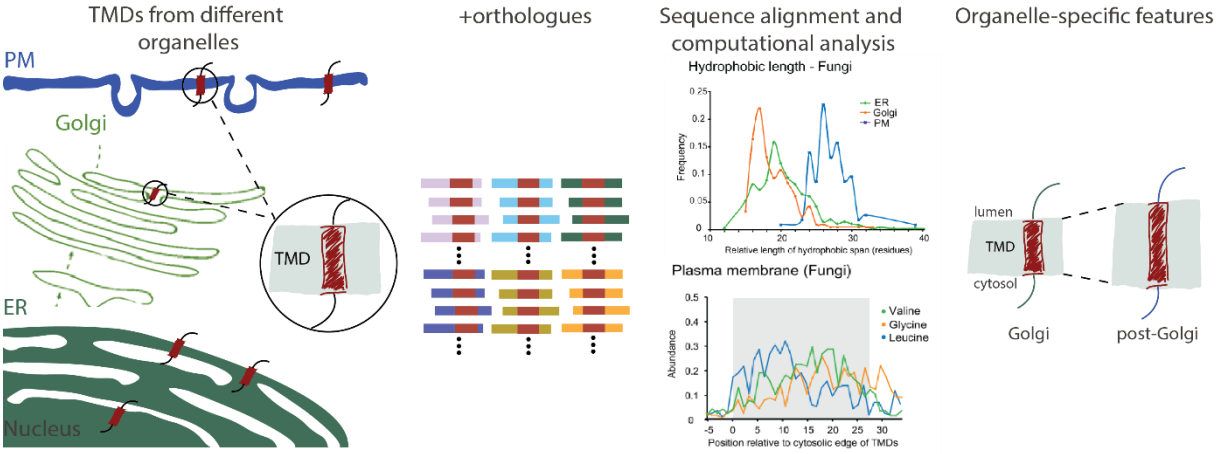

Figure 4. A comprehensive comparison of TMD's sequences from fungi and vertebrates transmembrane proteins [modified from (Sharpe et al., 2010)]. The initial datasets contained the sequences of all single TMD proteins from *S. cerevisiae* and *H. sapiens*. Literature searches and cross-referencing between databases were used to identify those proteins with a known organelle of residence and topology. Orthologues were added to the subset of the proteins with the single TMD and known localization. Afterwards, sequence alignment and computational analysis of the biophysical properties and AA composition were performed. The analysis revealed a dichotomy in TMD length between the early and late parts of the secretory pathway. TMDs from post-ER organelles show striking asymmetries in amino acid compositions across the bilayer that is linked to residue size and varies between organelles.

In 2010 a novel class of C-terminal signals was identified computationally. Previously the *S. cerevisiae* uncharacterized open reading frames (ORFs) YDL012C, YBR016W and YDR034W-B were shown to localize at the PM and were proposed to belong to the tail-anchored protein group (Beilharz et al., 2003; Huh et al., 2003). By using sensitive sequence profile analysis, Venancio and Aravind observed the enrichment in cysteine residues at the C-terminus for these PM-localized proteins and named it cysteine-rich transmembrane (TM) module (CYSTM) (Venancio and Aravind, 2010). This short cysteine domain with low hydrophobicity flanked with proline and glutamine at the N-terminal part is conserved across eukaryotes (Figure 5). Proteins with a CYSTM domain are reported to be involved in stress response and the deletion of the CYSTM domain disrupts PM localization of these proteins (Andreeva et al., 2017; Joshi et al., 2020; Mir and León, 2014; Pereira Mendes et al., 2021; Xu

et al., 2018). Whether or not the CYSTM is a TMD has not been confirmed experimentally. Previously identified PM-localized TA proteins Sso1, Sso2, Snc1 and Snc2 have long TMDs with very high hydrophobicity which correlates with the properties of the PM (Sharpe et al., 2010). It will be important to check if a short CYSTM module with very low hydrophobicity can anchor at the PM or CYSTM proteins are actually associated with the cell periphery via interactions with other PM proteins.

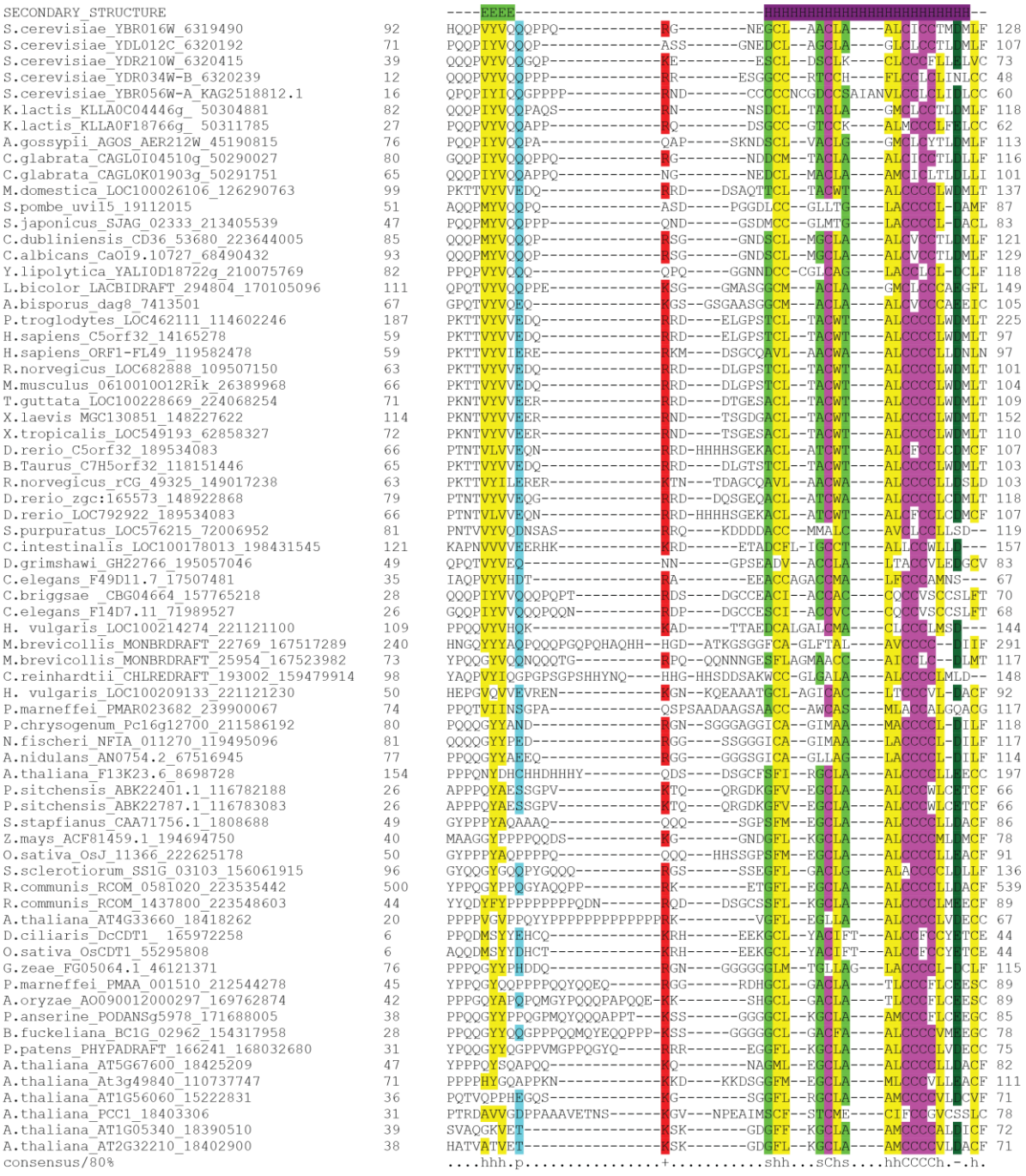

Figure 5. Multiple sequence alignment of the CYSTM domain superfamily [modified from (Venancio and Aravind, 2010)]. The columns were colored according to the consensus shown below the alignment (hydrophobic (h), polar (p), positively charged (+), small volume (s), negatively charged (-), and the predicted secondary structure is shown on top (E- extended confirmation (β-strand), H - helical). The numbers indicate the start (left) and end (right) of CYSTM module. The sequences are labeled using the species name, gene name, and GenBank gi number and sequence identifiers.

Different targeting and sorting pathways recognize specific TMD features but a systematic analysis of these features is required to dissect this group of localization signals. In the following chapters, I will describe what is known up to date about the targeting pathways for ER membrane proteins.

4.1.3 CO-TRANSLATIONAL TARGETING OF MEMBRANE PROTEINS TO THE ER

To ensure correct and balanced protein targeting to subcellular compartments multiple cascades of cytosolic chaperones, membrane receptors and protein translocons are working in parallel.

The first co-translationally acting targeting factor that was discovered was the SRP, which has an important role in targeting type I, type II and type III signal-anchored proteins into the ER (Figure 6).

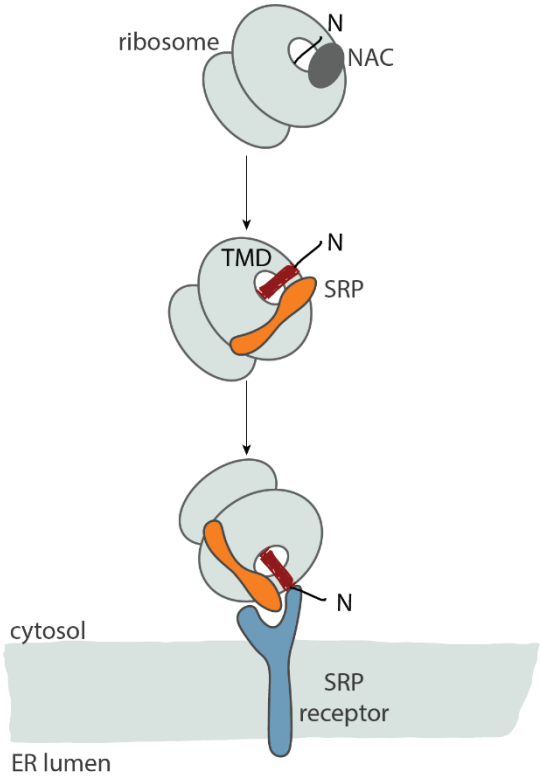

Figure 6. The co-translational SRP pathway. The nascent polypeptide-associated complex (NAC) prevents SRP from binding to the ribosome. Once a hydrophobic targeting sequence (signal peptide or TMD) is released from the ribosome, NAC is exchanged for SRP. Then SRP interacts with the SRP receptor at the ER membrane and brings the ribosome to the ER surface where the translation is coupled with the translocation into the ER lumen. N, amino terminus.

Once an N-terminal targeting signal emerges from the ribosomal exit tunnel, it is recognized by SRP that pauses translation and directs the ribosome-nascent chain complex (RNC) with the help of the cognate SRP receptor (SR) to the ER membrane (Gilmore et al., 1982; Rapoport, 2007; Walter et al., 1981; Zimmermann et al., 2011). Next, the ribosome and signal sequence are transferred to the Sec61 channel, protein synthesis resumes, and the nascent chain translocates into the ER lumen or ER membrane. The N-terminal targeting sequence can be a cleavable signal peptide of variable length or the first TMD within the polypeptide. SRP has a dual function - protecting the hydrophobic polypeptide during

translation and targeting the RNC to the ER. The targeting sequence has to be located in a distance of at least 65 amino acids from the polypeptide C-terminus as the site of targeting sequence recognition by SRP is positioned about 35 amino acids from the peptidyl transferase center inside the ribosome (Hegde and Keenan, 2021; Nyathi et al., 2013; Zhang and Shan, 2014). After recognition, SRP-mediated targeting to the ER takes about 5–7 seconds, allowing the synthesis of roughly 30 amino acids (Goder et al., 2000; Hegde and Keenan, 2021). If translation terminates during this time, co-translational targeting cannot happen. GTP hydrolysis by two universally conserved GTPases, the SRP54 - subunit of SRP and the other in the α- subunit of SRP receptor, provides the energy for substrate recognition by SRP, targeting the RNC to the SRP receptor and to recycle SRP to the cytosol (Hegde and Keenan, 2021).

Recently, by combining rapid auxin-induced SRP degradation with proximity-specific ribosome profiling Costa and colleagues showed that, in addition to the targeting of N-terminal signal peptides and N-terminal TMDs, SRP is generally essential for targeting TMDs regardless of their position relative to the amino terminus with the exception of TA proteins (Costa et al., 2018). Proteins with internal targeting signals universally depended on SRP and were susceptible to aberrant mitochondrial targeting in its absence. Together with work in bacteria, these findings suggest that the main function of SRP is to engage TMDs across the length of the nascent chain (Costa et al., 2018; Schibich et al., 2016). In addition, the discovery of an unexpected class of proteins that become mistargeted to mitochondria in the absence of SRP reveals an unanticipated role for SRP in maintaining the specificity of organelle targeting (Costa et al., 2018; Ingolia et al., 2019).

4.1.4 POST-TRANSLATIONAL TARGETING OF MEMBRANE PROTEINS TO THE ER

Membrane proteins with a single TMD at the very C-terminus are called TA proteins and are targeted to the ER post-translationally using either general or specialized chaperones in the cytosol (Hegde and Keenan, 2021).

TA proteins with TMDs of high hydrophobicity are typically targeted to the ER by the GET pathway in yeast or the TRC40/GET pathway in mammals (Denic, 2012; Guna and Hegde, 2018; Schuldiner et al., 2008; Wang et al., 2010) (Figure 7-a). In the yeast GET pathway, the TMD of a newly synthesized TA protein is captured by the chaperone Sgt2 (SGTA in mammals) near the ribosome surface. Next, the Get4/5 scaffolding complex recruits the Sgt2-TA complex via Get5, while Get4 recruits the central targeting factor, Get3, a homodimeric ATP-dependent chaperone. The Get3 dimer has a TMD-binding site in the middle that represents a ~ 35 Å long hydrophobic groove for an α-helix accommodation. An Sgt2-Get3 handover reaction is mediated by the Get4/5 complex. Get3 then targets the TA protein to the ER by a receptor composed of Get1 and Get2. At the membrane, Get1/2 disrupts the TA protein binding site in Get3, releasing the TA protein for insertion into the membrane. In the end, Get3 is recycled to the cytosol to initiate a new round of targeting (Hegde and Keenan, 2021; Mateja and Keenan, 2018).

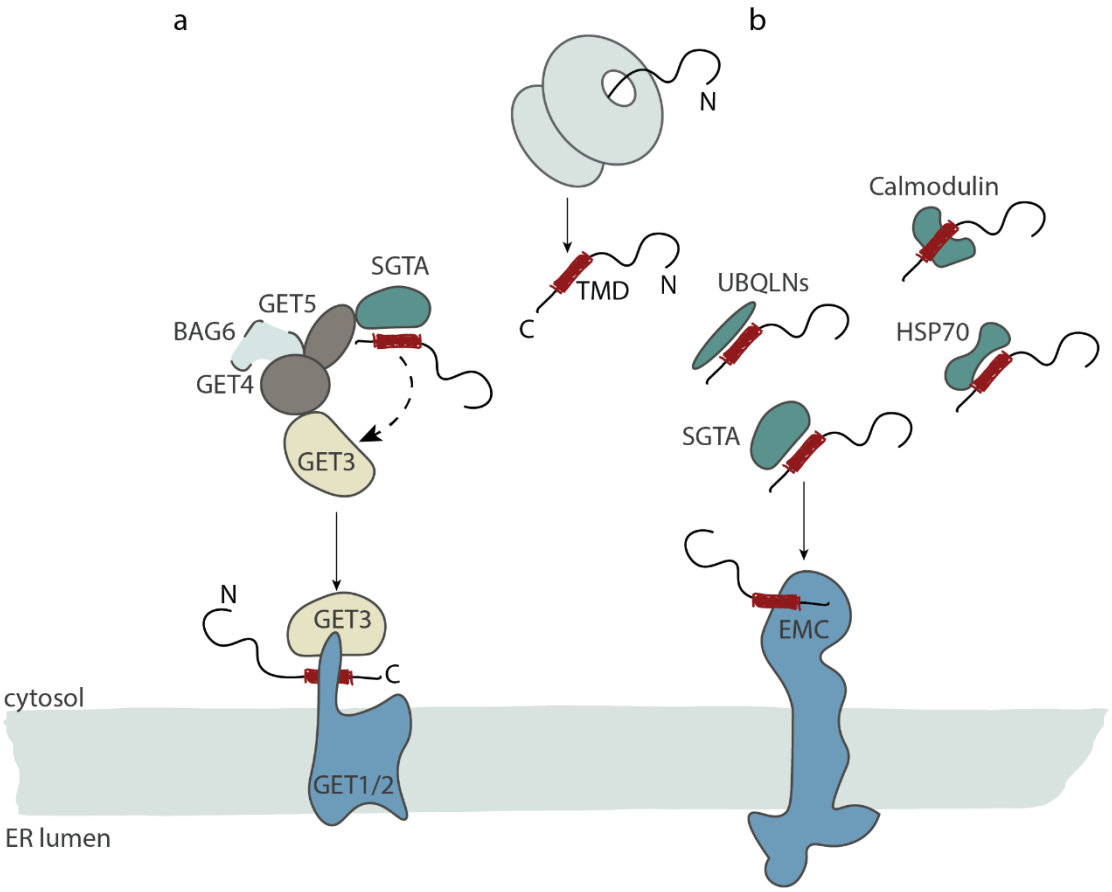

Figure 7. Post-translational targeting of TA proteins to the ER membrane. a - TA proteins with the TMDs of high hydrophobicity are captured by chaperone SGTA with the help of the GET4/5 complex (which in mammals also contains BAG6). From SGTA TA proteins are transferred to GET3. GET3 interacts with the GET1/2 complex. b - TMDs of low hydrophobicity are kept soluble in the cytosol by cycles of binding and release from any of several chaperones. Upon the release from the chaperon, the TMD engages the cytosolic domain of the EMC. C, carboxy terminus; N, amino terminus.

Recently, it has been demonstrated in mammalian cells that TA proteins are ubiquitinated as soon as they are synthesized in the cell (Culver and Mariappan, 2021; McQuown et al., 2021). The ubiquitinated TA proteins are directed to the targeting factor GET3 and inserted to the ER (Figure 8). After the insertion, the ER-associated deubiquitinases (DUBs), USP20 and USP33, remove ubiquitin chains from TA proteins (Culver and Mariappan, 2021). This finding suggests that TA protein biogenesis is a highly regulated process and that ubiquitination of nascent TA proteins could potentially protect their soluble domains from promiscuous interactions or prevent enzymatic activity until the proteins reach their target sites in the cell (McQuown et al., 2021). Another potential reason for ubiquitination of the nascent proteins could be that it may recruit ubiquitin-dependent chaperones and unfolding enzymes, such as p97 ATPase, to facilitate their folding and maturation (Culver et al., 2022). It has been shown that TA proteins carry a K48-linked polyubiquitin chain, which is typically a signal for proteasomal degradation. Interestingly, the TA proteins are not degraded, raising the question how they escape the degradation pathway (Culver and Mariappan, 2021). To complete the picture of TA protein biogenesis, it is necessary to identify the E3 ligases, DUBs, potential ubiquitin-dependent chaperones and unfolding enzymes.

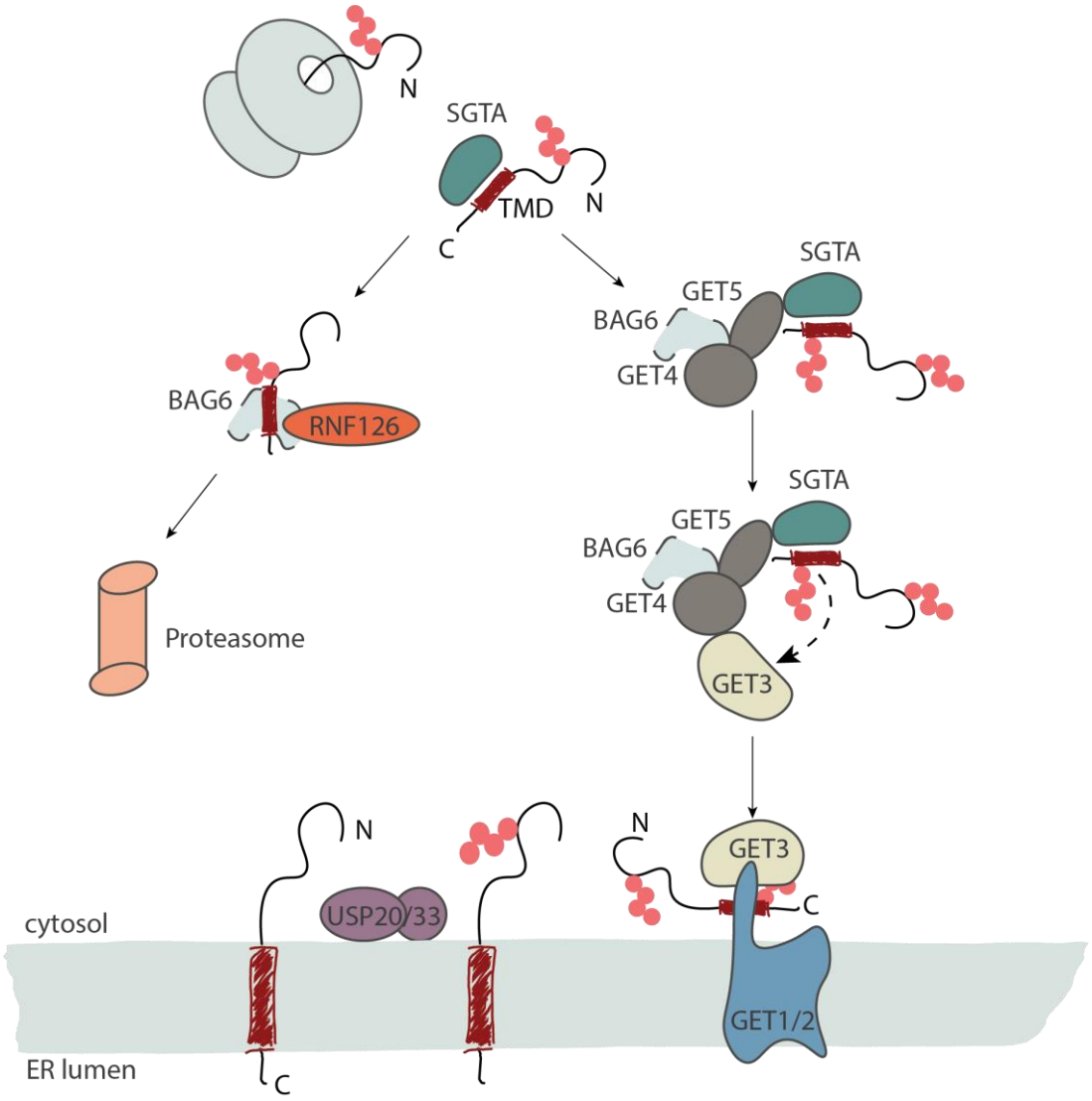

Figure 8. Ubiquitination and deubiquitination of TA proteins during protein biogenesis. TA proteins are ubiquitinated in the cytosol and captured by chaperone SGTA with the help of the GET4/5 complex (which in mammals also contains BAG6). TMDs with high hydrophobicity are transferred to GET3. GET3 interacts with the GET1/2 complex. TA protein is deubiquitinated at the ER membrane by USP20 and USP33. C, carboxy terminus; N, amino terminus.

TMDs of ER-localized TA proteins vary in length and hydrophobicity. GET/TRC40 substrates appear to have TMDs of high hydrophobicity. For a long time, it has been unclear how the ER-resident TA proteins with TMDs of moderate hydrophobicity are inserted into the membrane. Recent work in mammalian cells revealed that TA proteins with TMDs of moderate hydrophobicity are targeted to the ER by cytosolic chaperones and the EMC (Guna et al., 2018; Tian et al., 2019) (Figure 7-b). Such TMDs can also be captured by SGTA as it has a broad range of substrates. However, they are not transferred to GET3 (Lin et al., 2021). These TA proteins are kept soluble in the cytosol, with their aggregation prevented by dynamic binding and release from SGTA or other TMD-binding chaperones, including heat shock proteins such as HSP70, ubiquilin family proteins (UBQLNs) and calmodulin (Hegde and Keenan, 2021). Upon the release from chaperones, the TMD of these TA proteins is directly recognized by EMC at the ER membrane (Bai et al., 2020; Guna et al., 2018; Miller-Vedam et al., 2020; Pleiner et al., 2020).

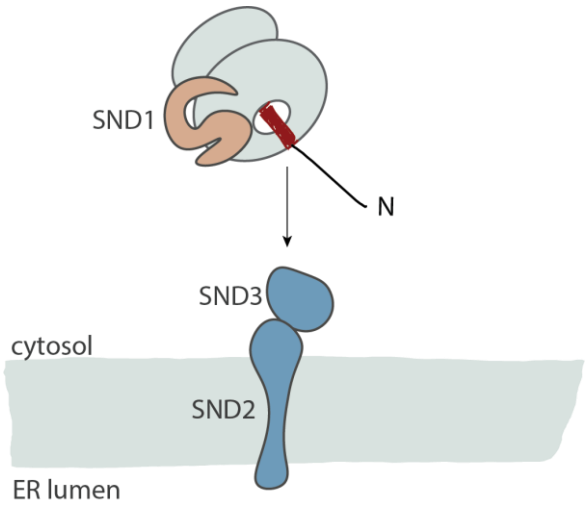

Figure 9. SRP-independent (SND) targeting of proteins to the ER membrane. SND1 binds to the ribosomes and then interacts with SND2 and SND3 receptors at the ER membrane. N, amino terminus.

In addition to the above-described pathways, there is a poorly understood SRP-independent (SND) targeting route to the ER (Figure 9). It has been observed that some SRP substrates can reach the ER even when SRP and GET pathways are compromised (Hegde and Keenan, 2021). In particular, it has been shown in yeast that glycosylphosphatidylinositol (GPI)-anchored proteins undergo SRP-independent targeting and translocation with the help of three proteins: Snd1, Snd2 (Tmem208 in humans) and Snd3 (Ast et al., 2013; Aviram et al., 2016). Snd1 interacts with the ribosome in the cytosol and then binds to ER-localized Snd2 and Snd3. SND-targeted proteins are translocated to the ER lumen through the Sec61 channel. The SND pathway has a broad substrate range and, probably, acts as a functional backup for the other targeting routes (Aviram et al., 2016; Haßdenteufel et al., 2017). It has been proposed that the SND is involved in the targeting of the transmembrane proteins with TMDs located far from N- and C-terminus. It is not clear if recognition of these substrates by SND happens co- or post-translationally.

4.2 CELLULAR QUALITY CONTROL MECHANISMS ENSURING PROTEIN HOMEOSTASIS

Correct localization of proteins ensures their functionality. The specificity and efficiency by which localization signals are recognized by their corresponding targeting factors are crucial for the fidelity of spatial organization of proteins within cells. Protein targeting is a complex process, which involves multiple overlapping pathways, and limits on specific binding might lead to inefficient targeting or mistargeting. Most protein targeting pathways face the additional problem that localization signals are degenerate and often depend on the biophysical properties (charge, hydrophobicity, helical content) rather than the amino acid sequence. In addition, properties of protein secondary and tertiary structures play a role as well. Furthermore, the localization signals for different destinations can share some degree of similarity, risking inappropriate recognition (especially in the case of TMDs) (Hegde and Zavodszky, 2019). Even under optimal conditions errors in protein targeting occur. Proteins inserted into the wrong compartment may become orphan due to lack of binding partners, or misfold and form aggregates due to changes in folding environment, which can potentially

lead to toxicity (Figure 10). This is underlined by the fact that just a single missense mutation in a localization signal can cause severe genetic disorders (Arnold et al., 1990; Cassanelli et al., 1998; Guo et al., 2014; Hung and Link, 2011; Hussain et al., 2013; Jarjanazi et al., 2007; Karaplis et al., 1995; Seppen et al., 1996). For example, it has been shown that the loss of a positive charge in the n region of preproinsulin signal sequence leads to impaired translocation across the ER membrane. Untranslocated preproinsulin accumulates in a juxtanuclear compartment distinct from the Golgi complex, induces the expression of heat shock protein 70 (HSP70), and promotes β cell death and diabetes (Guo et al., 2014). This is a clear example how the presence of protein in the wrong compartment leads to the toxicity.

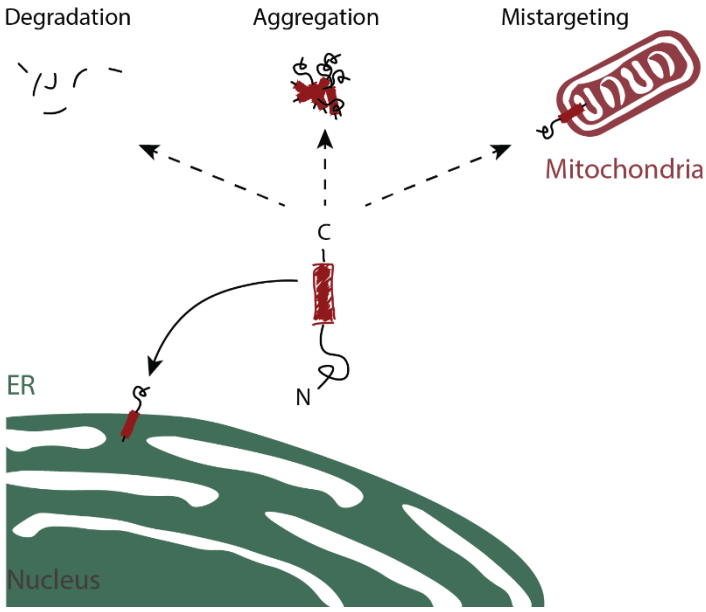

Figure 10. Multiple fates of the ER-resident TA protein. The TA protein can be inserted into the wrong compartment, form aggregates in the cytoplasm or be degraded.

To avoid the dramatic consequences of protein mislocalization cells have evolved protein quality control systems for maintaining proteome homeostasis through prevention, detection and removal of abnormal proteins. Translation, targeting, insertion, folding and assembly are monitored by a combination of cytosolic and membrane quality control factors.

4.2.1 QUALITY CONTROL FACTORS IN THE CYTOSOL

Due to the failure in targeting or insertion into the membrane, proteins can mislocalize to the cytosol. Some of the proteins in the wrong environment become misfolded. Their hydrophobic regions, which are normally buried inside the protein structure, start being exposed to the cytosol (Kong et al., 2021b). Many transmembrane proteins with hydrophobic TMDs are also at risk of degradation in the cytosolic environment, as well as the proteins with non-cleaved hydrophobic signal sequences. These hydrophobic regions are recognized by substrate-specific ubiquitin-protein ligases (E3) and targeted for degradation.

Misfolded proteins in the cytosol can be recognized by different chaperones that initiate chaperone-mediated refolding or in the case of irreparably damaged proteins (when the damage is caused by oxidation, carbonylation or nitrosylation) recruit E3 ligases

(Buchberger et al., 2010). Two chaperone families Hsp70 and Hsp90 are involved in chaperone-assisted degradation in mammalian cells. The engagement of Hsp70 or Hsp90 in a specific cellular process depends on its cooperation with a distinct set of co-chaperones (Arndt et al., 2007). One example is the carboxyl-terminus of Hsc70 interacting protein (CHIP) that interacts via a TPR domain with both Hsp70 and Hsp90 and possesses a U-box domain conferring ubiquitin ligase activity mediating substrates for degradation (Buchberger et al., 2010). CHIP initiates the ubiquitination of a wide range of Hsp70 and Hsp90 substrates, including steroid hormone receptors, transcription factors, and cystic fibrosis transmembrane conductance regulator (CFTR) (Arndt et al., 2007). In yeast the ER-resident carboxypeptidase Y and the vacuolar proteinase A become misfolded when they mislocalize to the cytosol. These misfolded proteins are recognized by the chaperones Ssa1 and Ydj1, which then initiate their degradation by recruiting the E3 ligase Ubr1 (Hegde and Zavodszky, 2019; Kong et al., 2021b).

Due to the failures in targeting pathways or limitations of targeting factors, proteins with hydrophobic signal sequences or TMDs can accumulate in the cytosol. Different quality control factors recognize exposed hydrophobic signal sequences and TMDs and facilitate substrate ubiquitination and degradation by the proteasome. It has been shown that the Bag6 complex can protect TA proteins from aggregation and that the E3 ubiquitin ligase RNF126 can be recruited to the N-terminal Ubl domain of Bag6 to ubiquitinate Bag6-associated substrates (Rodrigo-Brenni et al., 2014). The Bag6 complex has a dual function and is able to mediate the TA substrate transfer from SGTA to Get3 chaperone and to facilitate the efficient targeting of TA proteins to the ER in mammalian cells. Interestingly, not all ubiquitinated Bag6-associated substrates are sent for proteasomal degradation, which suggests the existence of other E3 ligases for a different ubiquitin modification (Culver and Mariappan, 2021). The ubiquilin family represents another group of cytosolic TMD-binding chaperones. It has been demonstrated that ubiquilins can bind TA and non-TA mitochondrial proteins and target them for proteasomal degradation (Itakura et al., 2016). However, the E3 ligases involved in this process remain unknown. In contrast, for the ER-destined transmembrane proteins, ubiquilins were shown to be involved in protein targeting (Hegde and Keenan, 2021). A possible explanation for such diverse functions lies in the mechanisms of ubiquilin interaction with the substrates. It is a dynamic process of constant binding and release cycles, which precludes membrane protein aggregation and, for a short period, allows the engagement of the targeting factors. If the interaction with the targeting factor does not happen, ubiquilins may recruit other E3 ligases and target membrane proteins for proteasomal degradation (Guna and Hegde, 2018).

4.2.2 QUALITY CONTROL FACTORS AT THE MEMBRANES

In yeast different E3 ligases such as Hrd1, Asi, Doa10 and Tul1 function directly at the membranes of their respective organelles. They target misfolded membrane proteins for proteasomal degradation and, additionally, control the abundance of some membrane proteins (Bays et al., 2000; Deak and Wolf, 2001; Swanson, Locher and Hochstrasser, 2001;

Reggiori and Pelham, 2002; Carvalho, Goder and Rapoport, 2006; Khmelinskii et al., 2014; Dederer et al., 2019) (Figure 11).

The polytopic E3 ubiquitin ligases Hrd1, Asi1 and Doa10 are the components of the ER-associated degradation (ERAD) pathway. Hrd1 contains a RING-H2 domain at the C-terminus that is required for Hrd1 activity. Hrd1 in complex with other proteins, including Ubc1 and Ubc7, the E2 ubiquitin-conjugating enzymes, acts at the ER membrane by recognizing exposed hydrophobic and charged residues of the misfolded proteins (Bays et al., 2000; Carvalho et al., 2006; Deak and Wolf, 2001). Another E3 ligase, Doa10 (MARCH6/TEB4 in mammals), a multipass membrane protein with a cytosolic RING finger domain, triggers the degradation of proteins with TMDs at N- and C-terminus and has been reported to be involved in the degradation of mistargeted mitochondrial proteins (Carvalho et al., 2006; Dederer et al., 2019). Misfolded or mislocalized proteins at the inner nuclear membrane are removed by the Asi1/3 complex (Foresti et al., 2014; Khmelinskii et al., 2014; Natarajan et al., 2020).

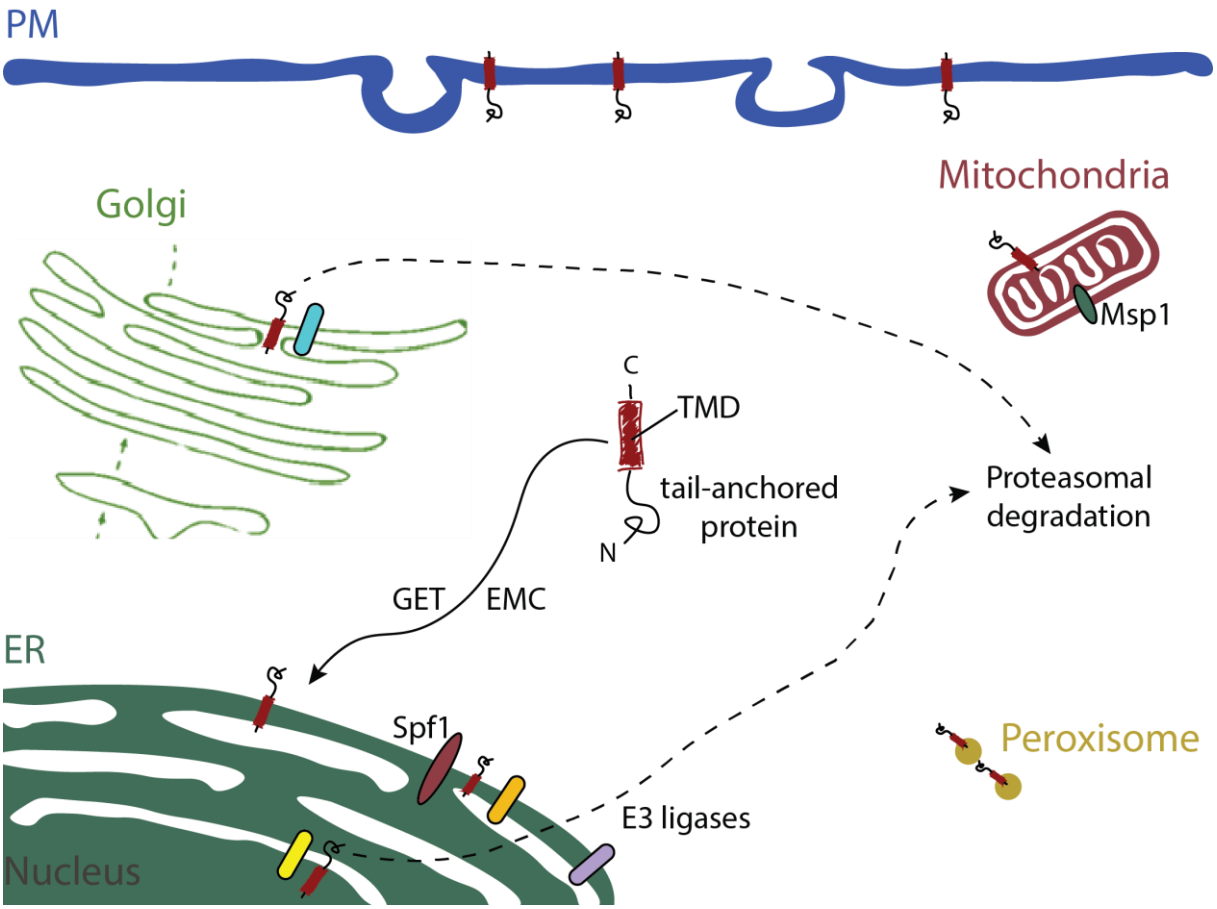

Figure 11. TA protein targeting and QC factors. TA proteins are targeted to the ER by GET/TRC40 or EMC pathway. In yeast different E3 ligases such as Hrd1, Asi1, Doa10 and Tul1 target misfolded transmembrane proteins for proteasomal degradation. Two protein dislocases, Msp1 and Spf1, remove mislocalized transmembrane proteins from the mitochondrial and ER membranes.

The Tul1 ubiquitin ligase functions in Golgi quality control. It is an integral Golgi membrane protein with a C-terminal RING domain that binds and functions with the E2 ubiquitin-conjugating enzyme Ubc4. Tul1 has been reported to target mutant SNARE proteins Pep12 and Tlg1 for degradation and, in later studies, a wider range of substrates was identified

(Reggiori and Pelham, 2002; Tong et al., 2014; Valdez-Taubas and Pelham, 2005). Recently Yang and colleagues showed that Tul1 performs protein regulation and quality control at the endosome and vacuole in addition to the Golgi (Yang et al., 2018).

A similar quality control system with E3 ligases Rsp5, Mdm30 and Ltn1 (Listerin in mammals) exists at the mitochondrial membrane and is called mitochondria-associated degradation (MAD) (Ravanelli et al., 2020). Both MAD and ERAD use the energy provided by the ATPase associated with diverse cellular activities (AAA) Cdc48 (p97 in mammals) (Dederer and Lemberg, 2021).

Two recently-discovered protein dislocases, Msp1 and Spf1 (ATAD1 and ATP13A1 in mammals), remove mislocalized proteins from the mitochondrial and ER membranes, respectively (Krumpe et al., 2012; Matsumoto et al., 2019; Mckenna et al., 2020; Okreglak and Walter, 2014; Qin et al., 2020) (Figure 11). Msp1 is a type I AAA-ATPase with an N-terminal extension encoding a TMD that removes mislocalized single-spanning membrane proteins from mitochondria and peroxisomes (Dederer and Lemberg, 2021). In Msp1-deficient cells Msp1 substrates such as Pex15, Fmp32 and Gem1 accumulate at mitochondria (Dederer et al., 2019; Okreglak and Walter, 2014; Weir et al., 2017). In addition to the extraction of mislocalized proteins, Msp1 dislocates orphan proteins, which failed to assemble into stable protein complexes (Dederer and Lemberg, 2021). Extracted proteins have different fates, some undergo proteasomal degradation, whereas others can be returned to the targeting cycle (Matsumoto et al., 2019). Spf1 is a P5-type ATPase with four domains: the N-domain, responsible for nucleotide binding; the catalytic core, referred to as the P-domain; an actuator A-domain; and the membrane-spanning M-domain responsible for ligand binding and transport across the membrane (Dederer and Lemberg, 2021; Mckenna et al., 2020). Spf1 deletion leads to the accumulation of mitochondrial proteins at the ER membrane and defects in lipid and sterol metabolism (Krumpe et al., 2012; Mckenna et al., 2020). Spf1 directly interacts with the TMD of mislocalized mitochondrial TMD proteins and extracts them into the cytosol potentially opening the route for retargeting (Mckenna et al., 2020).

The fate of a mislocalized TA protein after its dislocase-dependent extraction from a membrane for a long time remained a mystery. Is the TA protein directly recognized by the ubiquitin-proteasome system (UPS) in the cytosol and targeted for degradation or is it captured by a targeting factor and offered a second chance for correct targeting? Matsumoto and colleagues revealed that mislocalized TA proteins move from mitochondria to the ER in a manner strictly dependent on Msp1 expression (Matsumoto et al., 2019, 2022). Msp1 recognizes substrate TA proteins, extracts them from the mitochondrial membrane and facilitates their transfer to the ER. The ER-localized E3 ligase Doa10 ubiquitinates them with Ubc6 and Ubc7. Then, the AAA-ATPase Cdc48 extracts the ubiquitinated substrates from the ER membrane for proteasomal degradation in the cytosol (Matsumoto et al., 2019). Recently, using a time-lapse microscopy-based approach Matsumoto and colleagues demonstrated that this mitochondria-ER transfer of mislocalized TA proteins is disrupted upon knocking out the GET pathway components, which proves the involvement of the GET pathway in the transfer of mislocalized TA proteins from mitochondria to the ER (Matsumoto et al., 2022).

It seems that the tight collaboration of protein targeting and protein quality control systems is necessary to maintain protein homeostasis and correct targeting, but further studies are required for a better understanding of these processes.

4.3 PROTEIN MISLOCALIZATION AND DISEASES

Even though different quality control systems efficiently maintain protein homeostasis, it can be overloaded with the excess of mislocalized proteins upon the interruptions of the targeting pathways, environmental stress or mutations in localization signals.

Proteins that fail to be targeted to the correct compartment can have different fates – degradation, aggregation or localization to a wrong compartment (Figure 10). In addition to the loss of function or unregulated function, protein mislocalization increases the risk of inappropriate interactions with important cellular proteins such as protein quality control and nucleocytoplasmic transport factors, or damage to the integrity of membrane structures (Kong et al., 2021b; Lashuel et al., 2002; Milanesi et al., 2012; Olzscha et al., 2011; Park et al., 2013; Woerner et al., 2016).

It has been demonstrated that blocking of protein targeting pathways results in extensive protein aggregation and mislocalization (Costa et al., 2018; Guna and Hegde, 2018; Schuldiner et al., 2008). For example, upon the disruption of the GET pathway, TA proteins that are normally destined to ER form aggregates in the cytosol or get mislocalized to mitochondria (Costa et al., 2018; Schuldiner et al., 2008). Mutations in the human peroxisomal targeting factor *PEX7* impair the import of PTS2-containing proteins into peroxisomes, resulting in the lethal inherited disease, rhizomelic chondrodysplasia punctata (Lazarow, 2006).

The protein targeting machinery is sensitive to environmental perturbations. During ER stress protein import into the ER lumen is attenuated which leads to protein mislocalization and aggregation. However, the attenuation is substrate-selective and proteins that are required for stress alleviation, are not affected by the attenuation of protein import (Hegde and Zavodszky, 2019). Upon mitochondrial stress, protein import into mitochondria is attenuated as well. This initiates a stress response for stress alleviation but increases the load on the UPS, which has to deal with the mistargeted proteins in the cytosol. In this case, even a minor compromise of the degradation machinery might lead to disease.

Just a single missense mutation in a localization signal can cause severe genetic disorders (Arnold et al., 1990; Cassanelli et al., 1998; Guo et al., 2014; Hung and Link, 2011; Hussain et al., 2013; Jarjanazi et al., 2007; Karaplis et al., 1995; Seppen et al., 1996). Several examples of mutations in N-terminal signal peptides, NLSs and nucleolar localization signals (NoSs) are represented in Table 1 (Hung and Link, 2011; Jarjanazi et al., 2008).

Table 1. Genetic alterations in localization signal domains associated with human genetic disorders.

Gene	Disease	Mechanism	Mislocalization	Reference
CASR	Familial hypocalciuric-hypercalcemia; Neonatal hyperparahyroidism	Mutation of signal peptide	Impaired protein transportation to cellular membrane	(Pidasheva et al., 2005)
DSPP	Dentinogenesis Imperfecta	Mutation of signal peptide	Failure of protein translocation into the ER	(Rajpar et al., 2002)
GP9	Bernard-Soulier syndrome	Mutation of signal peptide	Improper signal peptide cleavage and reduced transportation to the cellular membrane	(Lanza et al., 2002)
SRY	Swyer syndrome	Mutation of NLS	Loss of nuclear localization	(McLane and Corbett, 2009)
SHOX	Léri–Weill dyschondrosteosis	Mutation of NLS	Cytoplasmic retention	(Sabherwal et al., 2004)
FOXP2	Speech–language disorder	Mutation of NLS	Loss of nuclear localization	(Mizutani et al., 2007)
RPS19	Diamond–Blackfan anemia	Mutation of NoS	Loss of nucleolar localization	(Da Costa et al., 2003)
AGT	Primary hyperoxaluria type 1	Polymorphism and/or mutation	Mitochondrial mislocalization	(Djordjevic et al., 2010)

These mutations result in impaired translocation, transportation or secretion of the protein leading to severe diseases. For instance, mutations in the N-terminal signal peptide of CASR, a cell surface glycoprotein expressed in the parathyroid gland and kidney lead to its reduced intracellular and plasma membrane abundance. The mislocalization in this case disrupts the maintenance of extracellular calcium homeostasis leading to familial hypocalciuric hypercalcemia (FHH) and neonatal severe hyperparathyroidism (NSHPT) (Pidasheva et al., 2005). Another example how the reduced presence of a protein in the correct compartment leads to the toxicity is the mutation of the signal peptide region in the bicistronic dentine sialophosphoprotein gene (*DSPP*). This mutation affects translocation of

the respective protein into the ER. This causes a loss of function of the dentine sialoprotein and the dentine phosphoprotein, which results in defective dentine biomineralization (Rajpar et al., 2002).

On the other hand, the presence of a protein in the wrong place can lead to the toxicity due to spontaneous gain-of-function. It has been reported that an altered localization of transcription factors such as NF- kB, activating transcription factor 2 (ATF2), cAMP response element-binding (CREB), p53, E2F transcription factor and NF-E2-related factor 2 (NRF2) might contribute to cell death commitment in several neurodegenerative diseases (Chu et al., 2007; Hung and Link, 2011).

4.4 MODERN METHODS TO STUDY VISUAL PHENOTYPES

High-content screening, which consists of high-throughput microscopy and automated image analysis, was developed to understand the interplay between protein sequence and visual phenotypes (Boutros et al., 2015; Mattiazzi Usaj et al., 2016).

A typical high-content screen is performed in an arrayed format such that different perturbations are tested in parallel but in separate wells in the array. This approach has yielded major insights into cellular organization and function in different model systems (Mattiazzi Usaj et al., 2016). But it is both costly and laborious, making it challenging to apply to a large number (10^4-10^6) of perturbations.

In contrast, screens performed in a pooled format are considerably more cost-effective and high-throughput. In this format, genetic perturbations are applied as a pool and cells with the desired phenotype are subsequently selected. For that, the phenotype of interest is typically linked to a selectable readout such as cell viability or expression of a fluorescent reporter, which can be used to collect the desired cell population using fluorescence-activated cell sorting. Targeted DNA or RNA sequencing is then used to identify the perturbations enriched in the selected cells.

Applications of pooled screening have increased dramatically over the last ten years. This is particularly prominent with genetic screens using CRISPR (clustered regularly-interspaced short palindromic repeats), which rely on Cas proteins and guide RNAs to interfere with gene function in high throughput (Hanna and Doench, 2020; Shalem et al., 2015), and with deep mutational scanning (DMS) experiments, designed to determine the functional consequences of sequence variation (Fowler and Fields, 2014). Both types of experiments take advantage of inexpensive synthesis of pooled oligonucleotide libraries, which are used to generate libraries of guide RNAs required for CRISPR screens or libraries of sequence variants that form the basis of a DMS experiment. However, applying pooled screening to phenotypes such as cell or organelle morphology, or even protein localization, has not been trivial due to the difficulty of linking visual phenotypes to an easily selectable readout.

Recently, various approaches that enable high-content pooled screens, which completely bypass the need to link visual phenotypes to a selectable readout, have been developed. Some of these rely on sequential hybridization of fluorescent oligonucleotide

probes (FISH) (Chen et al., 2015; Emanuel et al., 2017; Eng et al., 2019; Wang et al., 2019) or use *in situ* sequencing (Feldman et al., 2019; Lee et al., 2014) to read out genetic perturbations on single-cell level following imaging-based phenotyping. Another important advance was the development of an instrument capable of image-activated cell sorting (IACS), which combines high-throughput imaging with real-time image analysis and cell sorting (Nitta et al., 2018; Schraivogel et al., 2022). However, despite their potential, these methods are difficult to implement because they involve complex methodology or custom-built equipment.

Newly developed screening methods based on selective photoconversion partially address these limitations (Hasle et al., 2020; Kanfer et al., 2021; Yan et al., 2021). Thus, in Visual Cell Sorting approach (Hasle et al., 2020) cells are first engineered to express a photoconvertible fluorescent protein, such as Dendra2, that will serve as a phenotypic marker. Following automated imaging and image analysis, the microscope is directed to photoconvert Dendra2 from the green-fluorescent to the red-fluorescent state specifically in cells with the desired phenotype and the entire cell population is subsequently sorted according to the photoconversion state using fluorescence-activated cell sorting. Visual Cell Sorting makes use of readily available instrumentation and with a throughput of ~4 cells/s (although below the speed of ~100 cells/s in IACS (Nitta et al., 2018)). In principle up to ~10^4-10^5 perturbations can be analyzed with this method, making it suitable for large visual DMS experiments or visual CRISPR screens. However, the method is limited to four phenotypic bins and it is important to consider the dilution of the photoconverted signal over time due to cell division, especially in rapidly dividing organisms such as bacteria and yeast.

For some protein localization studies, the best option would be to have an array of variants of a localization signal (or different localization signals), but dealing with arrayed libraries is usually very laborious. The development of a straightforward screening approach in arrayed format would address the limitations of the number of phenotypic bins and allow using generated protein variants for downstream applications.

4.5 AIM OF THIS STUDY

The main goal of this study was to expand the high-content screening toolbox by developing iPAL (imaging pooled-to-arrayed libraries) scanning, a deep mutational scanning approach to dissect protein localization signals.

I applied iPAL scanning to investigate localization determinants of TA proteins. TA proteins carry a single α-helical TMD at their C-terminus, which enables them to localize to the ER, mitochondria, Golgi, NE, vacuole and PM. TMDs appear to have organelle-specific properties: hydrophobicity, length, charge and helical content (Beilharz et al., 2003; Borgese et al., 2019; Fry et al., 2021; Kaufmann et al., 2003; Keskin et al., 2017; Nguyen et al., 2018, 1993; Pedrazzini et al., 2000; Rao et al., 2016; Sharpe et al., 2010; De Silvestris et al., 1995). Additionally, it has been shown for a few individual proteins that mutations in the TMD can lead to degradation or aggregation and also to protein mislocalization (Figueiredo Costa et al., 2018; Keskin et al., 2017; Okreglak and Walter, 2014; Vitali et al., 2018). So far, a comprehensive analysis of TMD sequences has been conducted only computationally, therefore it was exciting to see – are these differences in TMD properties really important for localization *in vivo*?

In this study, for the first time in a high-throughput *in vivo* experiment, I demonstrated that the TMD alone acts as a localization signal and systematically dissected the features of the TMD important for targeting fidelity. In addition, by combining iPAL with genetic screens I identified several new substrates of membrane quality control components.

5 RESULTS

5.1 IPAL SCANNING IS REPRODUCIBLE AND EFFICIENT

How to better investigate the properties of localization signals? For some protein localization studies, the best option would be to have an array of variants of a localization signal (or different localization signals), but dealing with arrayed libraries is usually very laborious. The initial aim of my PhD project was to develop iPAL (imaging pooled-to-arrayed libraries) scanning, a deep mutational scanning approach to dissect protein localization signals, in which pooled libraries can be converted into verified arrayed libraries.

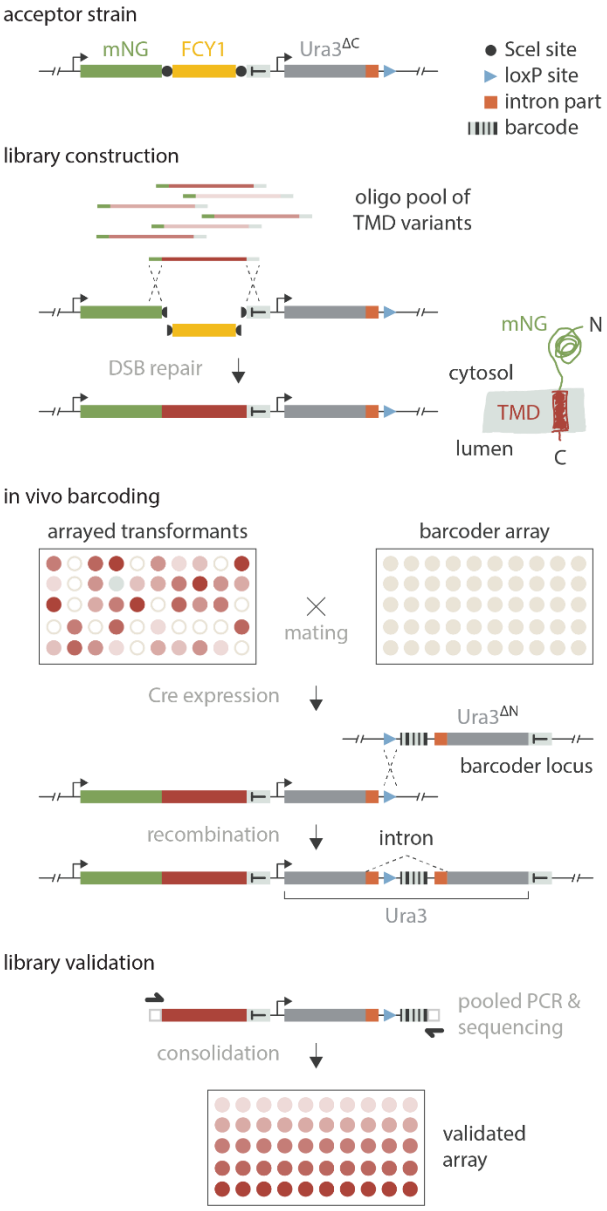

Figure 12. Experimental workflow of iPAL scanning. An acceptor strain has mNeonGreen (mNG) under the strong *GPD* promoter, followed by the selection marker *FCY1*, flanked by SceI cut sites, and a short synthetic terminator. Library construction: transformation of an acceptor strain with an oligo pool; transformants are selected on 5-FC plates for the loss of the *FCY1* marker; arraying of transformants in 1536 format. In vivo barcoding: mating with a *MATa* barcoding library, induction of Cre-recombination (both strains (*MATalpha* and *MATa*) contain loxP site for recombination); selection of diploid recombinants on SC-URA media (both strains (*MATalpha* and *MATa*) contain a part of *URA3* gene and a part of artificial intron and after mating functional *URA3* gene is formed as a result of splicing). Library validation: sequencing of barcoded variable regions; consolidation of the validated variants of interest from the starting haploid library into a new array; high-throughput fluorescence microscopy of mNG-TMD variants.

In iPAL, pooled libraries of protein variants with alterations in a potential localization signal are constructed by integrating oligo pools into an acceptor locus. For that, I created an acceptor strain, which has a fluorescent reporter (mNG), followed by the selection marker *FCY1*, flanked by SceI cut sites, and a terminator (Figure 12 - acceptor strain). Oligo pool, encoding localization signals, such as TMDs, is amplified by PCR with specific primers to extend homologous arms for genome integration. Then, the acceptor strain is transformed with an oligo pool and a plasmid encoding SceI endonuclease, which induces double-strand breaks (DSBs) at the integration locus. Then selection marker in the acceptor strain is efficiently replaced with the oligos encoding TMDs by homologous recombination (HR), leading to the expression of mNG-TMD fusions (Figure 12 - library construction). Using *in vivo* barcoding (Figure 12 - *in vivo* barcoding) and deep sequencing (Figure 12-library validation), pooled libraries are converted into verified arrayed libraries (Smith et al., 2017), followed by high-throughput fluorescence microscopy to determine the localization of each variant.

It has been shown that the TMDs of TA proteins are important for the correct localization of some TA proteins, but the TMD features crucial for targeting fidelity are not well defined. In order to define the localization determinants for all 59 yeast TA proteins, microscopy analysis of how their TMDs alone may affect localization is required. This is a good starting experiment for me to establish iPAL. Therefore, I decided to follow the above-described steps of iPAL to construct a library of TMDs from all 59 TA proteins (Figure 15) with the alterations of the TMD sequence and the flanking regions and fuse them with mNG to analyze their localization. The list of TA proteins used in this study contained TA proteins reported in the literature (Burri and Lithgow, 2004; Schuldiner et al., 2008) as well as several predicted potential TA proteins (Venancio and Aravind, 2010).

In the following subchapters, I will describe the steps of establishing iPAL in detail.

5.1.1 CREATING ACCEPTOR STRAIN

I have designed an acceptor strain that has mNG under the strong GPD promoter, followed by the selection marker *FCY1*, flanked by SceI cut sites, and a short synthetic terminator (Curran et al., 2015) (Figure 12 - acceptor strain). Fluorescent reporter (mNG) in the acceptor strain is separated from TMD integration locus with the short flexible linker to allow proper folding of the fluorophore. I have chosen mNG over sfGFP as mNG is the brightest monomeric green or yellow fluorescent protein yet described with exceptionally good performance as a fusion tag for traditional imaging (Shaner et al., 2013). Expressing mNG tag alone results in cytosolic localization and there is no evidence that free mNG is engaged in any translocation pathways. Another important feature of the acceptor strain is a short synthetic terminator T8 used in other studies to enhance protein expression (Curran et al., 2015). The advantages of using short synthetic terminators instead of commonly used *CYC1* or *TEF1* terminators are higher expression level and very short length (49 bp for the T8 terminator compared to 300 bp length for *CYC1* terminator). Thus, with T8 terminator I would be able to keep the DNA amplicon size under 900 bp, which is important for the efficient next-generation sequencing (NGS) reaction for the library validation (Figure 12-library validation).

Fcy1 is a cytosine deaminase, and it is required for *S. cerevisiae* to utilize cytosine as the sole nitrogen or sole pyrimidine source (Ear and Michnick, 2009; Hartzog et al., 2005). This marker is also counter-selectable. In the cell, cytosine deaminase converts 5-fluorocytosine (5-FC) to the toxic compound 5-fluorouracil (5-FU). This property allows me to select the transformants in which *FCY1* was replaced by an oligo encoding TMD variant. Flanking the *FCY1* cassette with SceI cut sites allows co-transforming the acceptor strain with SceI endonuclease, which induces DSBs. The replacement of the *FCY1* with the TMD oligo becomes more efficient, as DSBs are repaired by HR.

Figure 13. Optimizing a concentration of cytosine for FCY1 positive selection. SC-ADE/URA agar plates supplemented with 100, 50, 25, 15 and 10 mg/L cytosine. 1 – JHY650; 2 - SHA349; 3 - SHA345; 4 - #2797; 5 - #2836; 6 - #2849 (Table S 1). Strains with *FCY1* (1, 2, 3, 4, 6) and without *FCY1* (5 - *fcy1Δ*::HphMX) were streaked on cytosine plates to find out the optimal cytosine concentration for FCY1 positive selection.

To construct the desired acceptor strain, I first needed to integrate *FCY1* selection marker into the *ybr209wΔ* locus of the starting strain yEI0001. As a first step I optimized the working concentration of 5-FC and cytosine (Figure 13) for the selection of the *FCY1* cassette in my experiment. To determine cytosine concentration for *FCY1* positive selection, which is required for the acceptor strain construction, I streaked the strains with *FCY1* (1, 2, 3, 4, 6) and without *FCY1* (5) on SC-ADE/URA plates (prepared with yeast nitrogen base without amino acids and ammonium sulfate to exclude all nitrogen sources) supplemented with different amount of cytosine. 15 mg/L and 10 mg/L are optimal for *FCY1* positive selection as *fcy1Δ* strain (5) did not grow on these plates. Thus, to create the acceptor strain, I transformed the strain yEI0001 with GPDpr-mNeonGreen-(GA)x5-SceI$^{cut\ site}$-FCY1-SceI$^{cut\ site}$-T8term cassette and selected successful transformants on SC-ADE/URA+10 mg/L of cytosine. For *FCY1* negative selection, which is required for the replacement of the marker with the oligo encoding TMD variant, I tested 0.1 g/L, 0.5 g/L and 1 g/L 5-FC and SC plates supplemented with 1 g/L 5-FC showed the cleanest selection.

To test if *FCY1* in the acceptor strain can be efficiently replaced with the TMD oligos, I chose TMDs from 6 TA proteins with different localizations (Fis1 – mitochondria, Ubc6 – ER, Prm3 – NE, Tlg1 – Golgi, Nyv1 – vacuolar membrane, Sso1 – PM) (Figure 14-a). I transformed the acceptor strain with individual oligos encoding different TMDs and selected the transformants on 5-FC. The validation by Sanger sequencing showed that 90% of randomly picked clones carried the correct TMD sequence, which demonstrates that the transformation is working.

Microscopy of individually transformed strains with correct sequences showed that mNG-^{5}TMD^{5}(Fis1), mNG-^{5}TMD^{5}(Ubc6), mNG-^{5}TMD^{5}(Nyv1) fusion proteins showed expected

localization corresponding to full-length mNG-tagged proteins (Figure 14-a, b), whereas mNG-^{5}TMD^{5}(Sso1), mNG-^{5}TMD^{5}(Prm3), mNG-^{5}TMD^{5}(Tlg1) fusions localized to multiple compartments or differently from full-length mNG-tagged proteins (Figure 14-b). The differences in localization for mNG-^{5}TMD^{5}(Sso1), mNG-^{5}TMD^{5}(Prm3), mNG-^{5}TMD^{5}(Tlg1) fusions compared to full-length mNG-tagged proteins (Figure 16-b) could be caused by overexpression of synthetic construct, the effect of linker sequence (5 times GA repeat was used to separate fluorescent reporter and TMD) on TMD properties or it can be the evidence that TMD is not the only localization signal required for correct membrane targeting.

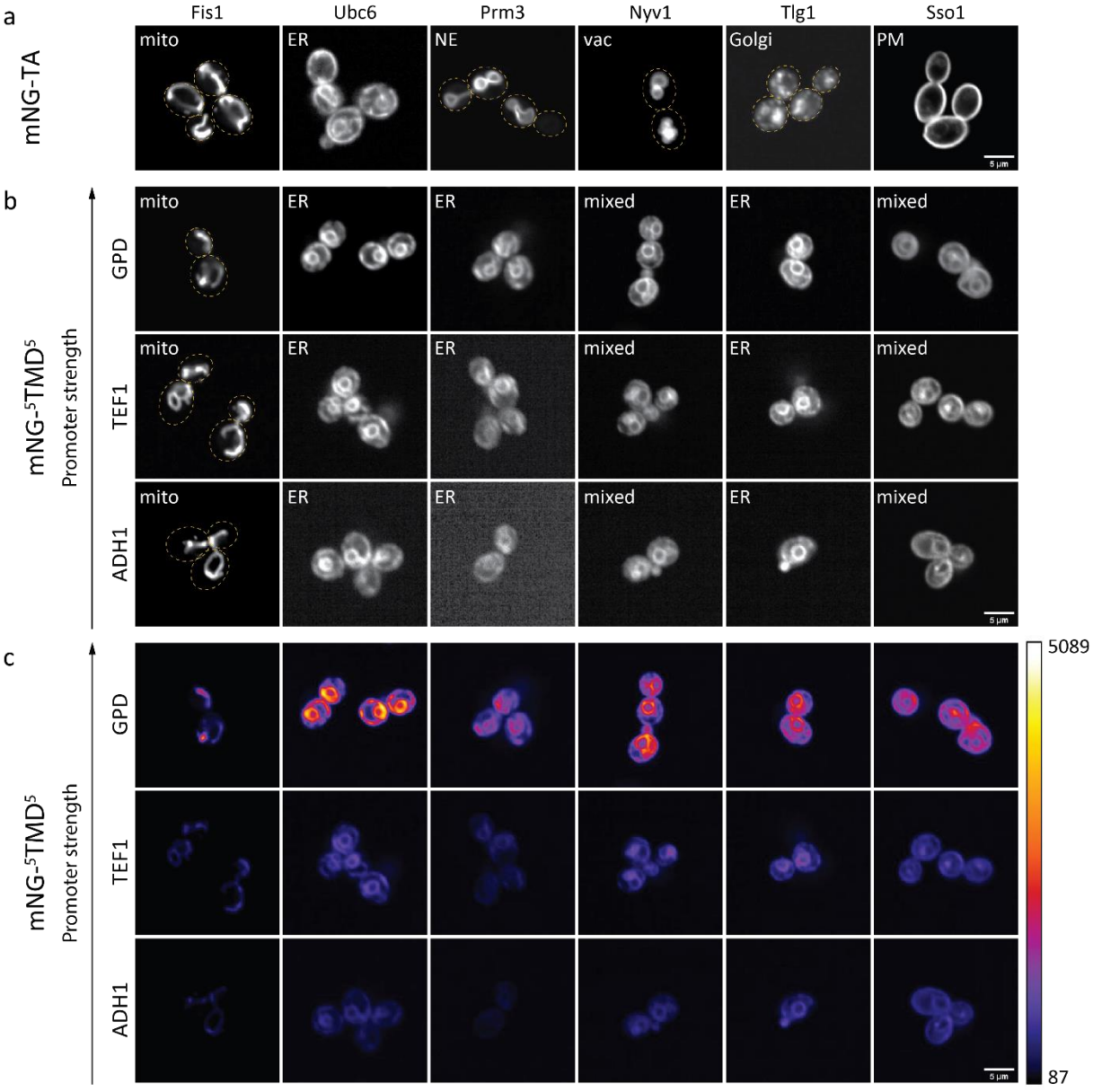

Figure 14. Testing different promoters for the acceptor strain.
a, b - Microscopy images of N-terminally mNG-tagged full-length TA proteins expressed under endogenous promoter (a) and corresponding mNG-TMD fusions expressed under different promoters (b). All strains were imaged with the same settings. Each image was scaled individually for optimal visualization. Localization class is indicated in the top left corner; the yellow dashed outlines indicate cell boundaries.
c - Microscopy images of mNG-TMD fusions expressed under different promoters. Fluorescence intensities are color-coded from black (lowest signal intensity 87) to white (highest signal intensity 5089). All strains were imaged with the same settings. For each mNG-TMD variant all images were scaled equally. Scale bar, 5 µm. ER, endoplasmic reticulum; mito, mitochondria; mixed, several compartments; PM, plasma membrane; vac, vacuole; NE, nuclear envelope.

To test whether overexpression has an impact on protein localization I created a set of strains with *TEF1* and *ADH*1 promoters, which are weaker than the *GPD* promoter. Microscopy analysis did not show any change in localization pattern (Figure 14-b), although expression levels differ significantly, meaning that overexpression conditions, in this case, do not affect the localization (Figure 14-c).

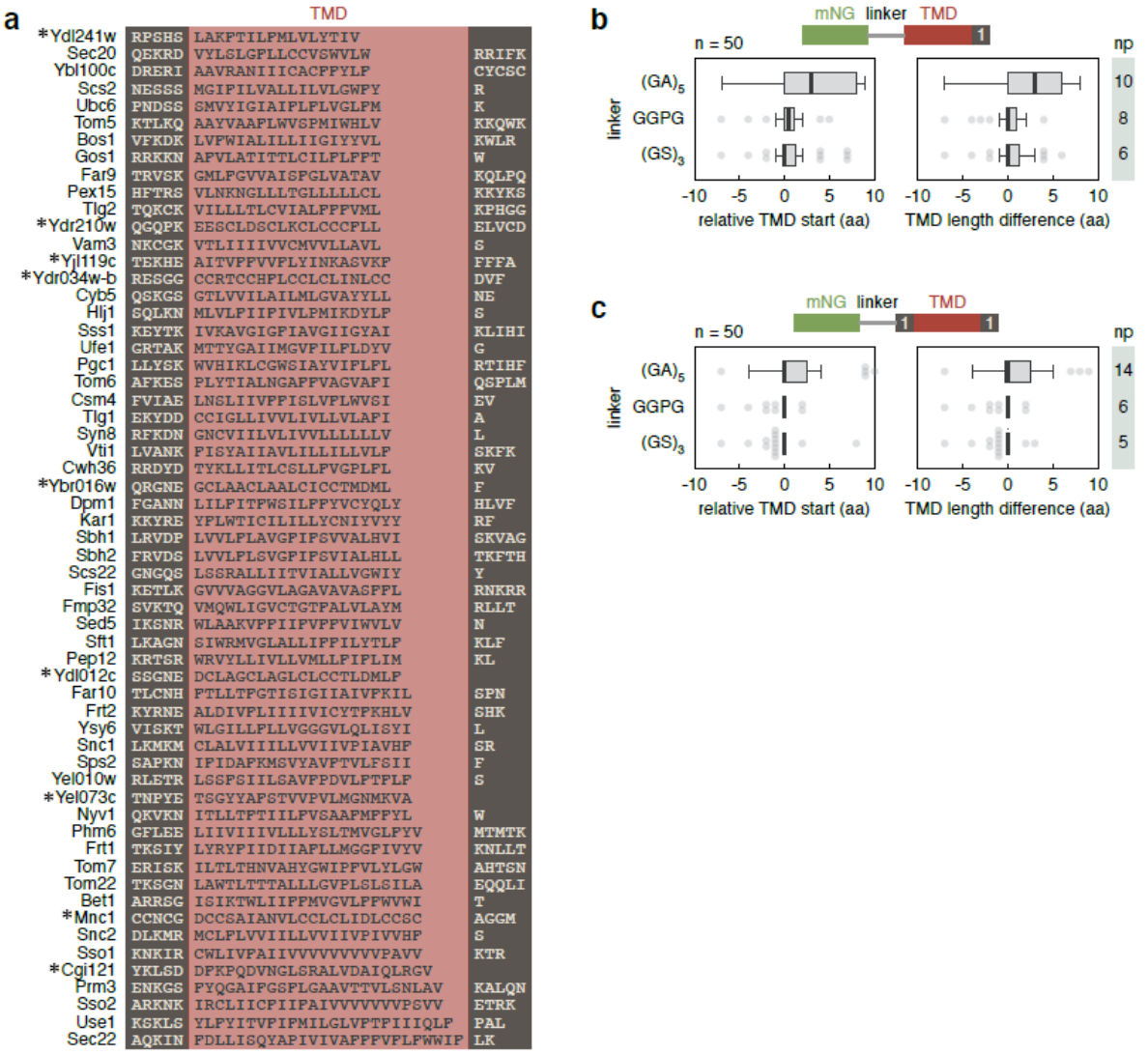

Figure 15. Linker selection for the mNG-TMD reporter.
a – TMDs and flanking residues of 59 yeast TA proteins, ordered by TMD length. TMD start and end positions were determined using full-length protein sequences and the Phobius predictor (Käll et al., 2004) except for 9 proteins marked with an asterisk. b, c – Influence of the linker between mNG and TMDs on predictions of TMD boundaries. Sequences of mNG-linker-TMD fusions using TMDs from (a), with the indicated number of native residues flanking the TMD (a single residue after the TMD (TMD^1, b) or a single residue before and after the TMD ($^1TMD^1$, c)) and three different linkers were analyzed using the Phobius predictor. Predicted TMD start and length are relative to predictions of full-length TA proteins in (a). Centerlines mark the medians, box limits indicate the 25th and 75th percentiles, and whiskers are the 95% confidence interval and any outliers are shown as circles. n – number of TMDs from (a) used in the analysis. np – number of mNG-linker-TMD fusions for which a TMD was not detected by the Phobius predictor.

To test if the GA linker sequence between reporter and TMD might affect the hydrophobic properties of TMD, I compared 50 yeast TA proteins' full-length sequences (Figure 15-a) (Burri and Lithgow, 2004) and corresponding mNG-TMD fusions using Phobius prediction tool (Käll et al., 2004), which can identify TMDs inside protein sequence. I have tested the TMD sequences with 1 flanking residue at the C-terminus (TMD^1) and with 1

flanking residue at N- and C-terminus (^{1}TMD^{1}). Relative start and length of TMD identified by Phobius for the mNG-(GA)x5-TMD^{1} (Figure 15-b) and mNG-(GA)x5-^{1}TMD^{1} (Figure 15-c) were very different from the expected start and length based on full-length protein set. The TMD predictions were lost for 10 mNG-(GA)x5-TMD^{1} variants (Figure 15-b) and for 14 mNG-(GA)x5-^{1}TMD^{1} variants (Figure 15-c). For 11 mNG-(GA)x5-^{1}TMD^{1} variants (Figure 15-c) the difference in the predicted TMD start was more than 3 AA. For example, for mNG-(GA)x5-^{1}TMD^{1}(Prm3) the TMD was not predicted and for mNG-(GA)x5-^{1}TMD^{1}(Scs2), mNG-(GA)x5-^{1}TMD^{1}(Hlj1), mNG-(GA)x5-^{1}TMD^{1}(Gos1) and mNG-(GA)x5-^{1}TMD^{1}(Snc1) the difference in the predicted TMD start was more than 9 AA, which is a significant change. Thus, there is a possibility that the differences in localization for mNG-^{5}TMD^{5}(Sso1), mNG-^{5}TMD^{5}(Prm3), mNG-^{5}TMD^{5}(Tlg1) fusions compared to full-length mNG-tagged proteins are caused by the effect of GA linker sequence.

In contrast, variants with GGPG sequence, used in other studies for the separation of multiple TMDs from each other (Hessa et al., 2005, 2007), showed a minor difference in the predictions for the relative start and length (Figure 15-b, c). The TMD predictions were lost for 8 mNG-GGPG-TMD^{1} variants (Figure 15-b) and for 6 mNG-GGPG-^{1}TMD^{1} variants (Figure 15-c). And only for 2 mNG-GGPG-^{1}TMD^{1} variants (Figure 15-c) the difference in the predicted TMD start was more than 3 AA. To minimize linker influence on TMDs properties I have tried to identify a more neutral linker computationally and performed the additional analysis for Myc-(GS)x3 linker sequence. The TMD predictions were lost only for 6 mNG-Myc-(GS)x3-TMD^{1} variants (Figure 15-b) and for 5 mNG-Myc-(GS)x3-^{1}TMD^{1} variants (Figure 15-c). Only for 2 mNG-Myc-(GS)x3-^{1}TMD^{1} variants (Figure 15-c) the difference in the predicted TMD start was more than 3 AA. Both mNG-GGPG-^{1}TMD^{1}(Prm3) and mNG-Myc-(GS)x3-^{1}TMD^{1}(Prm3) variants did not show any change in the TMD position prediction compared to the full-length Prm3 sequence. These results show that GGPG and Myc-(GS)x3 have more neutral properties than (GA)x5 linker and are more suitable for separating fluorescent reporter and TMD. For my further screens I exchanged (GA)x5 linker in the acceptor strain to Myc-(GS)x3 sequence and created a subset of the variants with GGPG sequence to compare Myc-(GS)x3 and GGPG linkers.

The final acceptor strain for iPAL yEI0038 (GPDpr-mNeonGreen-linker-SceI$^{cut\ site}$-FCY1-SceI$^{cut\ site}$-T8term) has mNG under the strong GPD promoter, followed by the selection marker FCY1 , flanked by SceI cut sites, and a short synthetic terminator T8 (Figure 12-acceptor strain). mNG in the acceptor strain is separated from TMD integration locus with the Myc-(GS)x3 linker to allow proper folding of the fluorophore.

5.1.2 OPTIMIZING TRANSFORMATION EFFICIENCY

To ensure the efficient integration of several hundreds of different TMDs into the acceptor strain, I optimized transformation efficiency. The selection marker *FCY1* in the acceptor strain is flanked by two SceI cut sites (Figure 12-acceptor strain). Upon the expression of SceI enzyme in the cells, DSBs are induced and the repair of the locus with homologous

fragments (Fis1 TMD with 45 bp homologous arms) is more efficient (Figure 16-b) compared to the experiment without SceI expression (Figure 16-a).

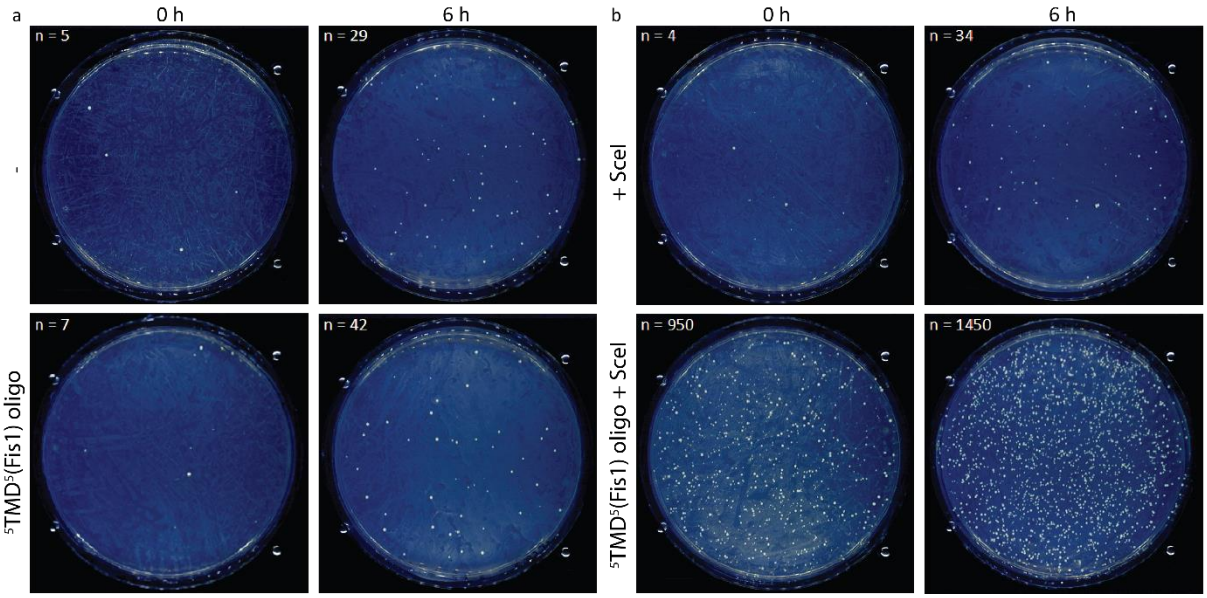

Figure 16. Optimizing transformation efficiency.
Testing transformation efficiency of the yEI0038 with ^{5}TMD^{5}(Fis1) oligo without the expression of SceI endonuclease (a) and with the expression of SceI endonuclease (b). Cells were plated on 5-FC media directly (0 h) or after recovery in YPD (6 h). n – number of colonies.

I amplified a trial oligo pool (containing just 6 oligos encoding different TMDs that previously showed membrane localization in fusion with mNG) by PCR with specific primers to extend homologous arms for genome integration from 18 and 22 bp to 45 and 45 bp. The minimum length of homologous arms of 45 bp is required for the efficient integration. Then I transformed the acceptor strain yEI0038 with the PCR product and the plasmid p413TEF-NLS-SCEI, encoding SceI enzyme, to induce DSBs. I plated the cells directly on 5-FC media or let them recover for 3 or 6 h in YPD (Figure 17-a). The recovery would allow the yeast to repair their cell walls and to completely lose already expressed Fcy1 protein. The results show that 6h recovery in YPD gives the highest amount of the transformants (1160 colonies) compared to 3 h recovery (953 colonies) and to direct plating (827 colonies). I also performed the transformation with freshly prepared competent cells (Figure 17-a) and competent cells stored at -80℃ (Figure 17-b). Plates with freshly prepared competent cells had a six-fold higher number of colonies (1160 colonies) than competent cells stored at -80℃ (175 colonies). Therefore, it is extremely important to use freshly prepared competent cells for library preparation.

Interestingly, there were always 30-70 colonies on the control plates (acceptor strain transformed with SceI enzyme without an oligo pool). These colonies most likely arise from spontaneous mutations in the *FCY1*, which make cytosine deaminize non-functional, and from non-homologous end joining (NHEJ) of the DSB. After 6 h recovery in YPD, the number of colonies on the transformation plates was fifteen-fold higher than on the control plates, but it is not clear how many of these colonies are correct transformants. Test microscopy of 96 randomly picked colonies showed that 70% of the transformants had membrane localization, whereas 30% had cytosolic localization, which confirms that the transformation works

efficiently as successful mNG-TMD transformants are expected to have membrane localization. Cytosolic localization might be present in the variants where *FCY1* was replaced with oligo which contained mutations, e.g., a premature stop-codon, or from the variants where *FCY1* got spontaneous mutations.

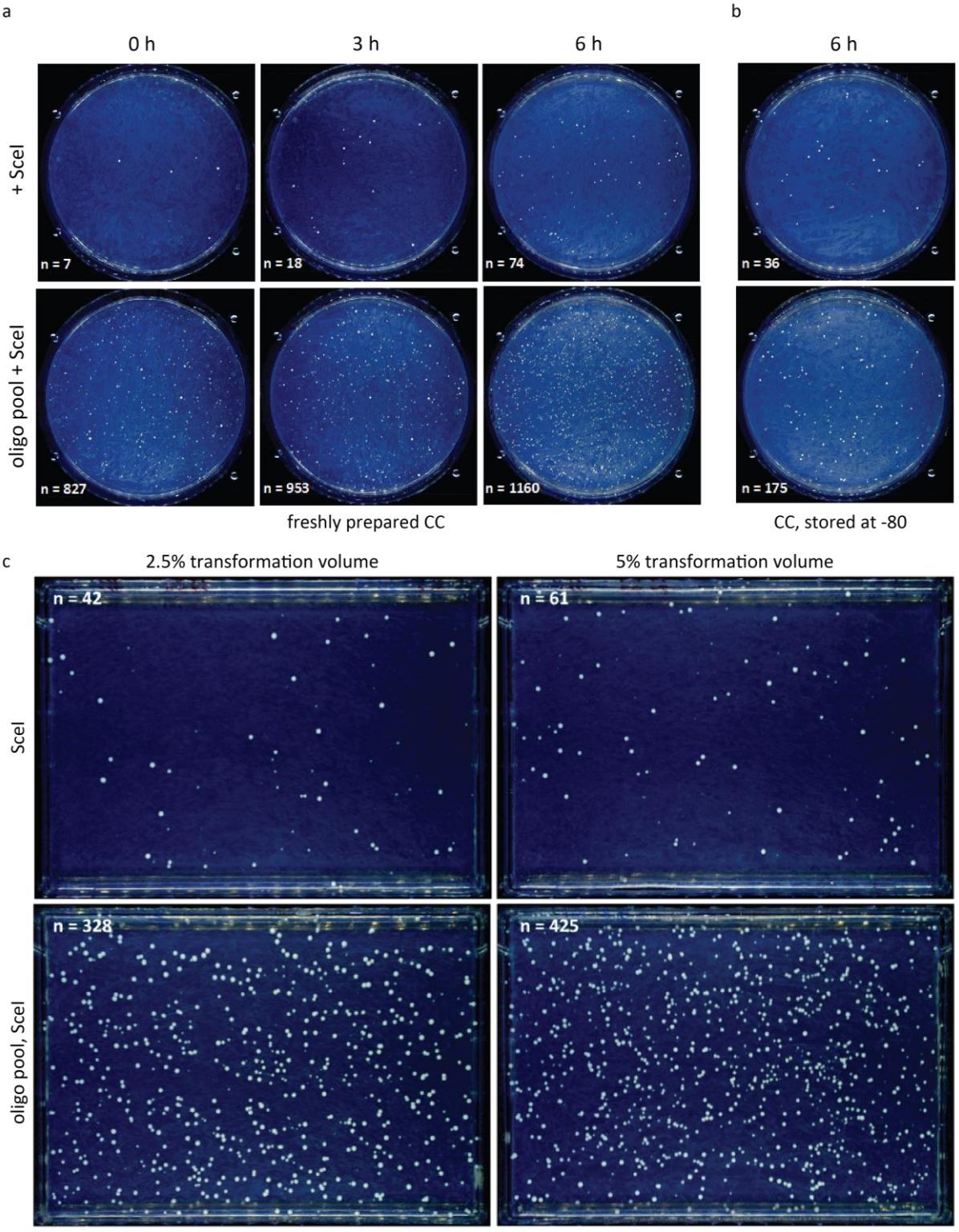

Figure 17. Optimizing transformation efficiency and colony picking procedure.
Testing transformation efficiency with freshly prepared competent cells (CC) (a) and with CC stored at -80 (b). Testing the optimal number of colonies per plate for the automated colony picking (c). n – number of colonies.

Taken together, transformation of freshly prepared CC with oligo pool and the plasmid encoding SceI endonuclease followed by 6 h recovery in YPD is efficient.

5.1.3 OPTIMIZING COLONY PICKING PROCEDURE

It is very important that the colony comes from a single cell to achieve reliable results in the experiment. Strains that arose from several cells show several TMDs matched to the same barcode at the sequencing step and heterogeneous localization at the microscopy step. I have plated different amounts of cells to find out the optimal density of the colonies for colony picking using an automated colony picking robot (Figure 17-c). Plating 2.5% of the transformation volume gives the optimal number of colonies per plate (250-300). Test microscopy of 96 re-arrayed colonies showed that 98% of the strains had homogeneous localization, which confirms that each colony is clonal.

5.1.4 OPTIMIZING BARCODING PROCEDURE

To identify the position of the TMD variants on the well plate, I used *in vivo* barcoding procedure previously described in the study from Smith et al. with slight modifications (Smith et al., 2017). The barcoder array consists of 1536 strains of *MATa* mating type. Each strain carries a unique 26-bp barcode associated with the strain position on the plate and a loxP site for recombination. The strains in the barcoder array (*MATa*) and the re-arrayed transformants (*MATalpha*) also contain Cre recombinase encoded under inducible *GAL* promoter. Mating both arrays and inducing the expression of Cre recombinase by switching the carbon source from glucose to galactose results in recombination in the diploid cells. Cre recombinase recognizes loxP sites encoded in both strains and physically links variable TMD region and the barcode. Selection of diploid recombinants is possible due to the presence of a part of URA3 gene and a part of artificial intron in both strains. As a result of splicing, functional Ura3 protein is produced and recombinants can be selected on the medium lacking uracil. Then the recombinants strains can be pooled together, with the TMD-barcode part amplified by PCR and validated by sequencing.

For barcoding, the array of the transformants is crossed with the barcoder array on YPD. Next, the expression of Cre recombinase is induced by re-plating the strains on YP+Raf/Gal plates (raffinose and galactose are used as a carbon source). Selection of diploid recombinants is performed on SC-URA (synthetic complete media without uracil). I have tested the possibility of plating the strains on YP+Raf/Gal and then on SC-URA or directly on SC-URA+Raf/Gal (Figure 18-a). Plating on SC-URA+Raf/Gal resulted in the slower growth of the strains. Plating only once on SC-URA after YP+Raf/Gal led to the presence of colonies of different sizes, which is not optimal for pooling the strains together because the proportion of each colony in the pool should be similar and different colony sizes will lead to a bias. To achieve maximum efficiency of the barcoding and have a similar colony size, I plated diploid strains on YP+Raf/Gal twice and performed the selection of the recombinants on SC-URA twice (Figure 18-b). Plating the strains on SC-URA twice also helped to get rid of the traces from the strains that failed the recombination step.

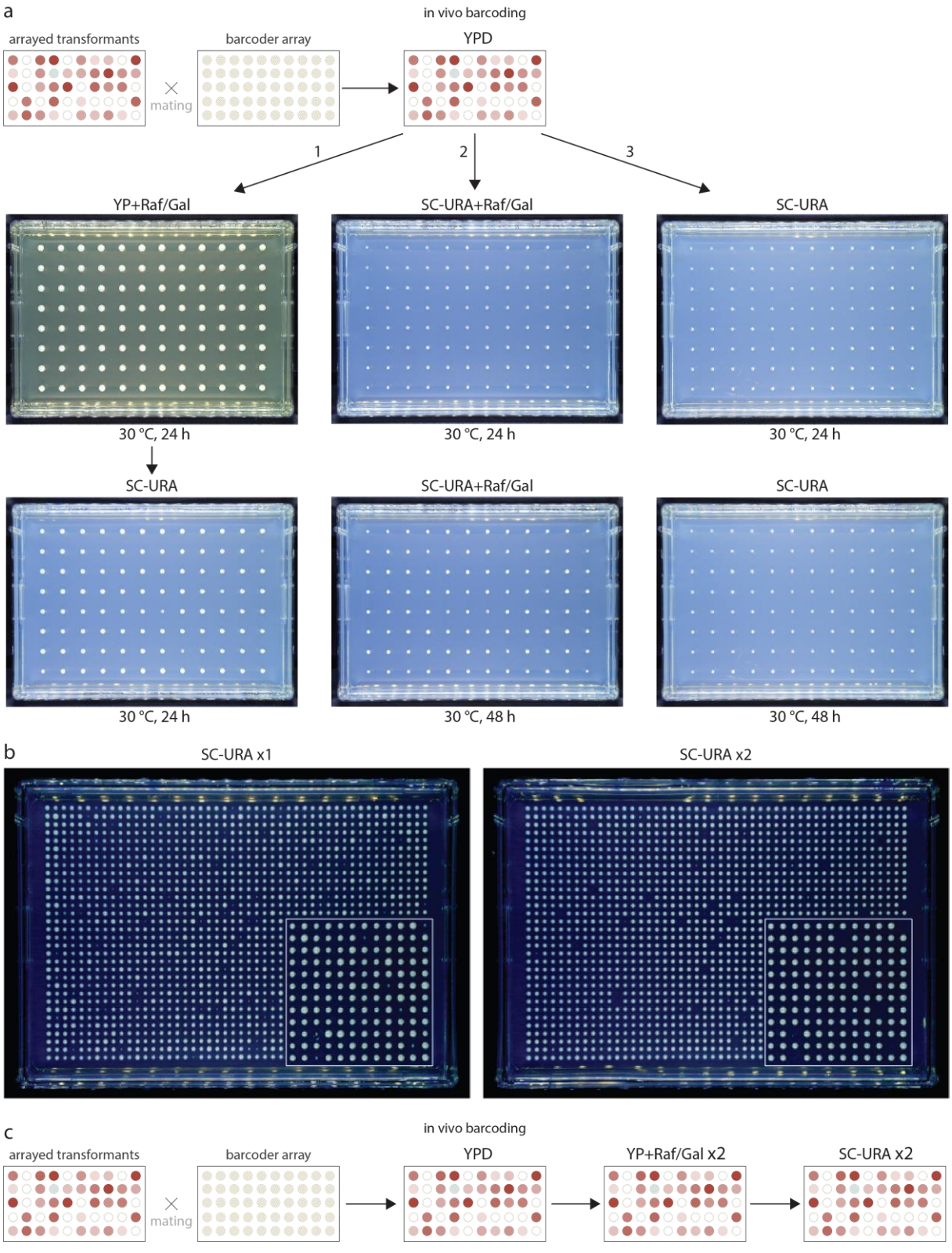

Figure 18. Optimizing barcoding procedure.
a - Testing the media for *in vivo* barcoding. 1 – Diploids were stamped on YP+Raf/Gal, incubated for 24h for recombination induction, and then stamped on SC-URA and incubated for 24h for the selection of the recombinants. 2 - Diploids were stamped on SC-URA+Raf/Gal and incubated for 24h and 48h to test combined recombination and recombinants selection. 3 – Diploids were stamped directly on SC-URA and incubated for 24h and 48h (negative control, recombination was not induced and colonies did not grow on the selection media).
b - Plating the strains on SC-URA twice makes the colony size more even.
c – Final barcoding procedure used for iPAL: diploids were pinned and grown on Raf/Gal plates in two rounds. Successful recombination events were selected in two rounds by growth on SC-URA plates. Raf – raffinose; Gal – galactose.

The following *in vivo* barcoding procedure was used for iPAL (Figure 18-c): stamping diploids on YP+Raf/Gal twice and selecting the recombinants on SC-URA twice.

5.1.5 LIBRARY PREPARATION FOR SEQUENCING

To cover all oligo variants and have biological replicates I picked a number of colonies equal to approximately 20x fold size of the oligo pool (e.g., to cover the pool of 384 TMD variants with different flanking residues I picked 7815 colonies). The barcoder array consists of 1536 unique barcodes, therefore I prepared several libraries for sequencing with different indexes and pooled them together for the NGS step. The recombinants were stamped on SC-URA agar plates in 1536 format. In order to pool all recombinants with 1536 unique barcodes from one plate to make one library, the recombinants were washed with SC media, and then total genomic DNA was purified from ~150 million cells.

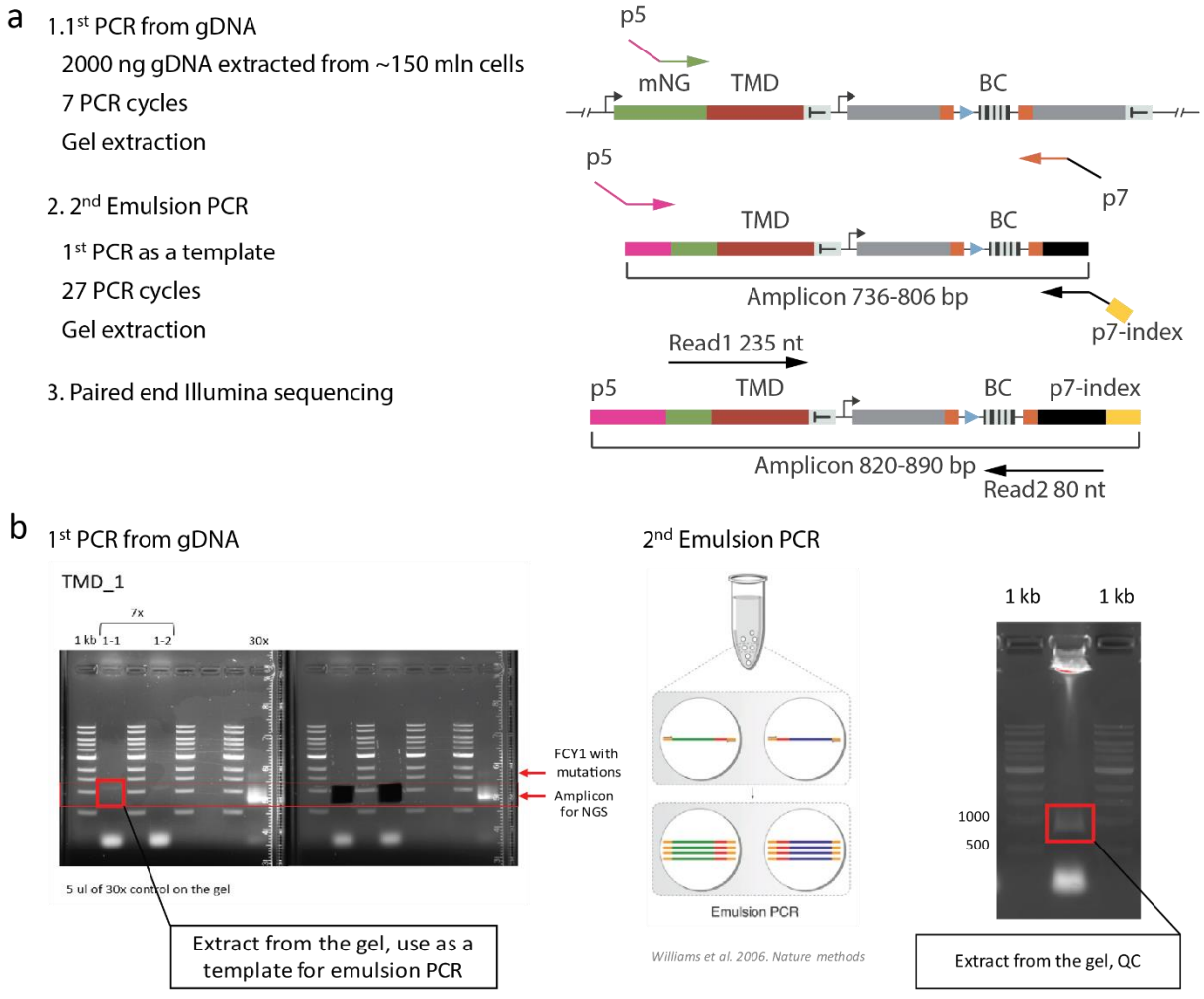

Figure 19. Library preparation for sequencing.
a – A scheme of the DNA amplicon size for the first and second PCR and primer binding regions. During the first PCR, TMD-barcode (BC) region is amplified with the primers containing parts of p5 and p7 adapters for Illumina sequencing. During the second PCR, p5 and p7 adapter regions are extended and p7-indexes are added to mark different libraries. Second PCR is performed in the emulsion of oil mixture and aqueous phase to avoid template switching.
b – DNA fragments for gel extractions from the first and second PCR. TMD-barcode amplicon has a size around 800-900 bp, and *FCY1*-barcode amplicon has a size of 1500 bp.

The genomic region of interest containing variable TMDs and 26 bp unique barcodes was PCR-amplified from 2 µg gDNA in 2 steps PCR (Figure 19-a,b), using emulsion PCR at the

second step to avoid template switching (Williams et al., 2006). I have tested 0.5, 1, 2 and 4 µg of gDNA as a template and 2 µg gave the best yield.

DNA amplicon sequencing is limited to PCR products of 1000 bp, because longer amplicons will not be able to cluster on the flow cell efficiently. In my experiment, the PCR gives two products – 900 bp TMD-barcode amplicon and 1500 bp *FCY1*-barcode amplicon in the case when the *FCY1* picked up spontaneous mutations and was not replaced by the oligo (Figure 19-b). 1500 bp product can bind to the flow cell but it is unable to form the bridges due to the large size. It is important to get rid of the leftover *FCY1* marker to avoid blocking the flow cell with the unwanted product. It is not possible to separate 900 and 1500 bp amplicons using beads purification, therefore I have decided to perform gel extraction at both PCR steps (Figure 19-b). The quality of the prepared libraries was assessed using Bioanalyzer. Several libraries (different plates) were pooled together in equal molarity and sequenced using MiSeq Micro Flow Cell (FC).

5.1.6 LIBRARY DESIGN AND PROOF OF PRINCIPLE EXPERIMENT

It has been reported that the context around the TMD is important for the correct localization of some TA proteins. Therefore, I used iPAL to construct a library (Figure 20-a) of TMDs from all 59 TA proteins (Figure 15) flanked by one or five amino acids (AA) from the native protein sequence and fused them with mNG with a short Myc-(GS)x3 linker or short linker and GGPG sequence. Myc-(GS)x3 and GGPG linkers did not show a dramatic change in TMD predictions compared to (GA)x5 linker (Figure 15-b ,c) and I wanted to compare Myc-(GS)x3 and GGPG linkers, used in other studies for the separation of multiple TMDs from each other, in the experiment (Hessa et al., 2005, 2007). The analysis of the localization of the TMD variants with different number of flanking residues separated from mNG with two different linkers is a good proof of principle of iPAL and it allows defining localization determinants for all yeast TA proteins.

Paired-end DNA sequencing was performed using MiSeq Micro FC. 235 cycles were used to sequence the exogenous DNA in the forward direction, and 80 sequencing cycles were used to sequence the barcode in the reverse direction. Sequencing reads were binned according to short (i.e. six nucleotide) indexing barcodes included in the reverse primers of the second PCR step (index was unique to each plate of diploid recombinants), and, then, by the 26-bp barcode (unique to each barcoder strain and, thus, colony position in the arrayed plate of recombinants) (done by Jorge Bouças, Max Planck Institute for Biology of Ageing). These steps required sequences to perfectly match the designed barcodes (all other sequences were excluded from further analysis). The TMD sequence in each colony was then identified as the most commonly observed sequence (at least 65% frequency) between the common priming regions, in each set of binned reads. 98% of the barcodes were identified for each library. The same barcodes on the different plates had different TMDs, which indicates that there was no cross contamination during the library preparation step. Only the variants with more than 100 reads were selected for microscopy experiment.

98% of TMD variants from the starting pool were present in the library, and 93% of the variants had a biological replicate, which demonstrates that barcoding and sequencing procedures are working well. 20 randomly picked variants were analyzed by Sanger sequencing and had barcode and TMD sequences identical to NGS results.

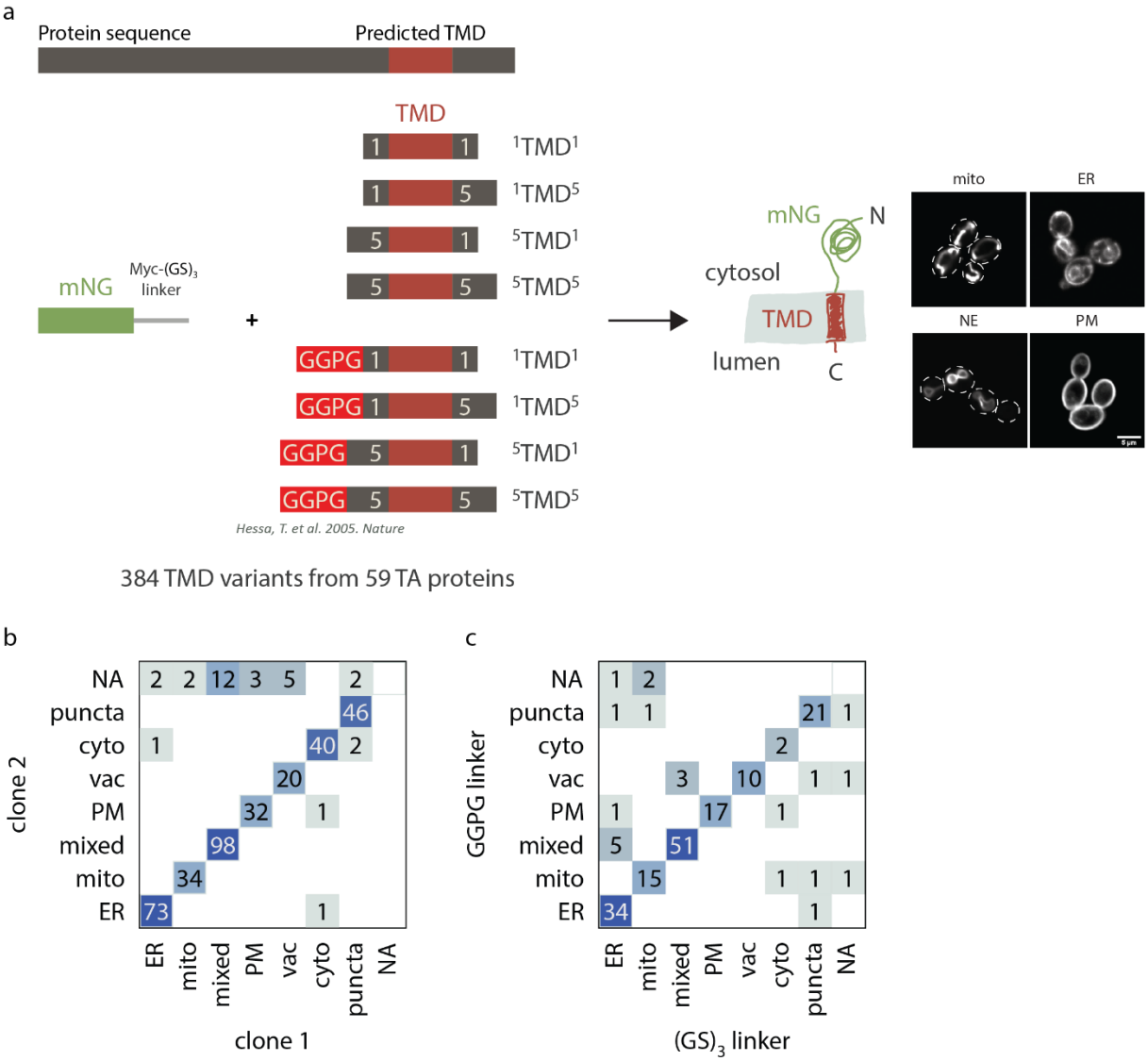

Figure 20. Library design and proof of principle experiment.
a - The TMD sequences from 59 yeast TA proteins, with the indicated number of native residues flanking the TMD (a single residue before and after the TMD (^{1}TMD^1), a single residue before and 5 residues after the TMD (^{1}TMD^5), 5 residues before and a single residue after the TMD (^{5}TMD^1), 5 residues before and 5 residues after the TMD (^{5}TMD^5) without and with GGPG separating sequence (Hessa et al., 2005, 2007)), used for the integration into the mNG-TMD acceptor strain with Myc-(GS)x3 linker. The white dashed outlines indicate cell boundaries; scale bar, 5 µm.
b - Comparison of observed subcellular localizations of mNG-TMD fusions from two biological replicates.
c - Comparison of observed subcellular localizations of mNG-TMD variants with two different linkers. ER, endoplasmic reticulum; mito, mitochondria; mixed, several compartments; PM, plasma membrane; vac, vacuole; cyto, cytosol; NA – mNG-TMD variants, which were not identified.

After the library validation by sequencing, I consolidated the correct TMD variants into the final array and performed high-throughput fluorescence microscopy to determine the subcellular localization of the variants. Microscopy analysis of 696 strains revealed that 98% of the biological replicates for TMD variant show the same localization (Figure 20-b). 98% of the strains had homogeneous localization, which shows that iPAL protocol is working well, each transformant is clonal and there is no cross-contamination during the preparation for

microscopy. TMD variants with Myc-(GS)x3 and Myc-(GS)x3-GGPG linkers showed very similar localization (Figure 20-c). 90% of the variants show the same localization, 5% of the variants localize at the connected compartments (e.g., ER and mixed or vacuole and mixed) and only 5% show completely different localization, which might be related to the linker effect for the particular variant or spontaneous mutations within the reporter sequence. This demonstrates that Myc-(GS)x3 linker is neutral and does not affect TMD properties for most of the variants, as well as that the iPAL approach is reproducible and reliable.

90% of TMD variants showed certain localization pattern (membrane compartments or puncta) and 10% of the variants showed cytosolic localization. Half of the variants that showed cytosolic localization were not predicted to have a TMD, therefore, there is a possibility that they actually do not have a TMD and are not TA proteins. 68% of the mNG-TMD fusions anchored at the ER, mitochondria, PM, vacuole or several membrane compartments simultaneously (mixed) and 22% of the mNG-TMD fusions showed puncta localization. These results prove that TMD alone can act as a localization signal, but detailed analysis is required to understand the correlation between TMD sequence properties and the localization.

5.1.7 ESTABLISHING AUTOMATED IMAGE ANALYSIS

As an outcome of high-throughput fluorescence microscopy in my proof of principle screen, I had hundreds of images with different subcellular localization, which I annotated manually. Increasing the number of TMD variants to investigate localization determinants of TA proteins in detail will result in thousands of microscopic images making the manual annotation extremely time-consuming and erroneous.

Therefore, we have decided to use machine-learning approach for data classification and segmentation, which would allow assessing two parameters – protein localization and abundance (by quantifying signal intensities). To perform automated classification and segmentation I created the strain yEI0064 (Table S 1) expressing cytosolic tagBFP2 for cell segmentation and the ER marker Sec61-mScarlet-I (mSc-I). Haploid mNG-TMD libraries were crossed with the strain yEI0064 and imaging was performed in diploid cells. This allowed me to introduce cell segmentation marker for automated classification and signal intensity quantification into mNG-TMD strains and directly validate the localization of ER and mixed (when the protein localizes at several compartments simultaneously, typically ER, Golgi and vacuole) categories. In collaboration with and , we have established an automated classification algorithm to classify 5 localization categories: ER, mitochondria, PM, vacuole and mixed (Figure 21-a). NE, Golgi and peroxisomes were underrepresented in our dataset and we did not include these categories in the pipeline. Cytosol and punctate images had a very high variety of signal intensities and were difficult to predict. Therefore, the image analysis was performed in the following way: first, I checked the images manually to identify cytosol and puncta, excluded them from the dataset and run the predictions for the rest of the categories. Then I combined the data from the predictor with my manual annotation for cytosol, puncta, NE, peroxisome and Golgi (Figure 21-b).

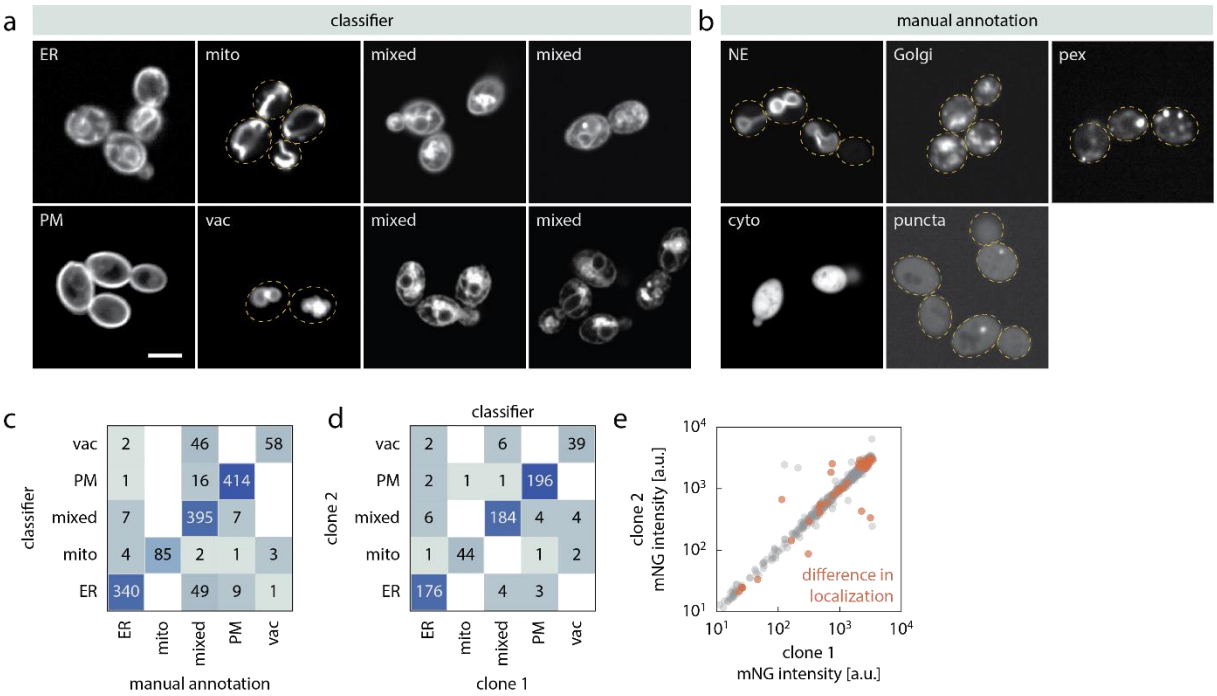

Figure 21. Establishing automated image analysis.
Example images of localization categories predicted by the classifier (a) and annotated manually (b). Localization class is indicated in the top left corner; the yellow dashed outlines indicate cell boundaries; scale bar, 5 µm.
c – Comparison of subcellular localizations of mNG-TMD variants predicted by the classifier and annotated manually.
d - Comparison of localization predictions of mNG-TMD fusions from two biological replicates. ER, endoplasmic reticulum; mito, mitochondria; mixed, several compartments; PM, plasma membrane; vac, vacuole; cyto, cytosol.
e - The correlation between protein abundance of the mNG-TMD variants from two biological replicates. Orange: mNG-TMD variants with the difference in localization. a.u - arbitrary units.

As we needed to make sure that the predictor is reliable, later, I performed manual annotation of all the images in my screens and compared it with the predictions (Figure 21-c). The results show that the predictor can identify the localization of five categories with high accuracy (96% for the ER class, 100% for the mitochondria class, 78% for mixed class, 96% for PM class and 94% for vacuole class). Automated classification pipeline often confused mixed category, which typically contains ER, Golgi and vacuole, with single ER or vacuole categories, because in some cases one of these three compartments was dominant. In this case this type of misclassification is not an error and the results can be used.

In total 98% of TMD variants were identified in my screens, and 95% of the identified variants had a biological replicate. Microscopy analysis of total of 2633 strains revealed that 98% of the biological replicates for the TMD variant show the same localization. Figure 22-c shows the correlation between classifier predictions and manual annotation for the images from the flanking residues screen and screen with different TMD extensions, as a subset of the images from these two experiments was used to train the predictor. Microscopy analysis of the biological replicates for these TMD variants shows that 95% of the biological replicates have the same localization predictions (Figure 21-d). Biological replicates also have a high correlation of the mNG intensities (97% of biological replicates show very similar mNG intensity) (Figure 21-e).

5.1.8 MICROSCOPY VALIDATION WITH CO-LOCALIZATION MARKERS

To validate microscopy data I have created a set of co-localization markers, expressing cytosolic tagBFP2 for cell segmentation and different marker proteins tagged with mSc-I at the C-terminus (ER marker Sec61-mSc-I, mitochondrial marker Cox4-mSc-I, Golgi marker Anp1-mSc-I, peroxisomal marker Pex3-mSc-I) (Figure 22-a, b). Co-localization assay with these markers allows the validation of all localization categories predicted by the classifier.

First, all haploid mNG-TMD libraries were crossed with the strain yEI0064 (Table S 1) and imaging was performed in diploid cells to quantify signal intensity and perform automated classification and directly validate the localization of two most abundant localization categories (ER and mixed). 96% of the strains classified as ER colocalized with the ER marker.

Second, strains with mitochondria and puncta localization classes were consolidated into a new array and crossed to the strains yEI0063, yEI0072 and yEI0073 (Table S 1) to validate mitochondrial localization and distinguish between peroxisomal, Golgi and protein aggregates for the strains with puncta localization. All strains classified as mitochondria colocalized with the mitochondrial marker. Interestingly, puncta localization observed for mNG-TMD fusions did not co-localize with Golgi or peroxisome markers (Figure 22-b) suggesting it being protein foci containing, possibly, protein aggregates. 2% of the strains classified as puncta colocalized with the mitochondrial marker.

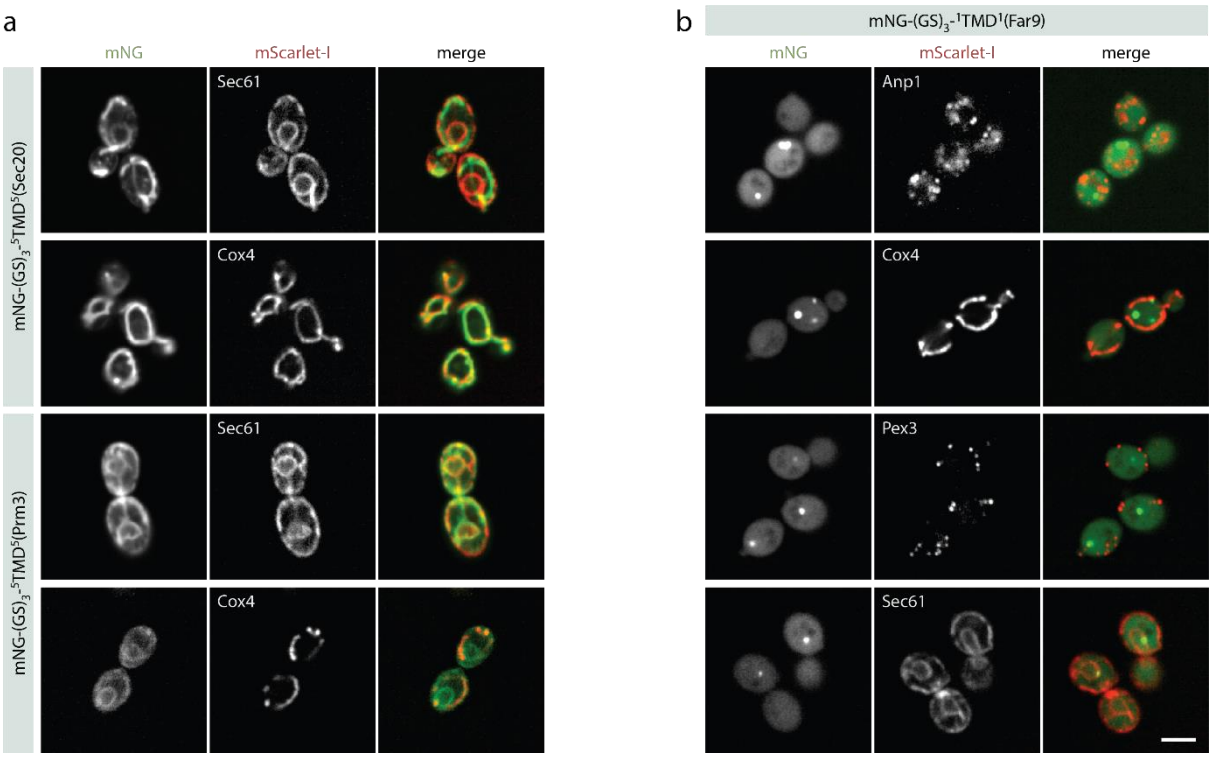

Figure 22. Validation of subcellular localization of mNG-TMD fusions using colocalization markers.
a – Colocalization of mNG-^{5}TMD^{5}(Sec20) variant with mitochondrial marker Cox4-mScarletI; colocalization of mNG-^{5}TMD^{5}(Prm3) variant with ER marker Sec61-mScarletI.
b – mNG-^{5}TMD^{5}(Far9) variant showed puncta localization and did not colocalize with Golgi marker Anp1-mScarletI or peroxisomal marker Pex3-mScarletI. Scale bar, 5 µm.

To sum up the results of section 5.1, the transformation of the acceptor strain is efficient and gives enough colonies to cover 98% of the variants in the pool and have biological

replicates for more than 90% of the variants. The amount of failed transformants is about 30%. 98% of the barcodes were identified for each library by NGS. The same barcodes on the different plates had different TMDs, which indicates that there was no cross contamination during the library preparation step. Microscopy showed that 98% of the strains had homogeneous localization, which confirms that each strain is clonal. Manual annotation of the microscopy images confirmed that the automated image classifier can identify the localization of five categories with high accuracy. 96% of the strains classified as ER colocalized with the ER marker and 100% of the strains classified as mitochondria colocalized with the mitochondrial marker, which proves that the image classification is accurate. Taken together, these results demonstrate that iPAL is reproducible and efficient.

5.2 MNG-TMD FUSIONS ARE RECOGNIZED BY THE TARGETING AND QC SYSTEMS OF TA PROTEINS

To make sure that mNG-TMD fusions behave as TA proteins in the cell, I first generated a set of full-length TA proteins tagged with mNG that can be used for comparison. I picked 50 strains encoding different sfGFP-tagged tail-anchored proteins from the N-SWAT-sfGFP collection (Yofe et al., 2016) (each ORF of *S. cerevisiae* is tagged with GFP at the N-terminus under the *NOP1* promoter) and swapped the sfGFP tag with mNG under the control of endogenous promoter or GPD promoter. The 9 ORFs that were not present in the N-SWAT-sfGFP collection were tagged individually and added to the collections (it was not technically possible to tag Yel010w and Ycl007c). The localizations of the mNG-tagged proteins expressed from endogenous promoter (mNG-ORF) or from GPD promoter (GPD-mNG-ORF) are listed in Table S 3.

Using iPAL I constructed a library of TMDs from 59 TA proteins (Figure 15) flanked by one or five amino acids (AA) from the native protein sequence and fused them with mNG with a short linker. From this library, I took a subset of the 53 TMDs flanked by five amino acids (AA) from the native protein sequence (mNG-^{5}TMD^{5}) that showed membrane localization to compare with full-length TA proteins tagged with mNG (mNG-TA). Both mNG-^{5}TMD^{5} and mNG-TA were under the control of GPD promoter. I did not include 6 mNG-TMD variants in this experiment as they showed cytosolic localization.

To test if mNG-^{5}TMD^{5} fusions are recognized by the protein targeting and QC systems in a similar way as mNG-TA proteins, I performed a genetic screen. Two known targeting pathways for TA protein targeting are the GET pathway and the EMC-dependent pathway, of which Get3 and Emc3 are the major components respectively (Figure 7) (Guna et al., 2018; Hegde and Keenan, 2011). Different E3 ligases such as Tul1, Hrd1, Asi1 and Doa10 can control the abundance of membrane proteins and target them for proteasomal degradation (Figure 11) (Bays et al., 2000; Carvalho et al., 2006; Deak and Wolf, 2001; Dederer et al., 2019; Foresti et al., 2014; Khmelinskii et al., 2014; Natarajan et al., 2020; Reggiori and Pelham, 2002; Swanson et al., 2001). Also two protein dislocases Msp1 and Spf1 remove mislocalized proteins from the mitochondrial and ER membranes (Figure 11) (Krumpe et al., 2012; Matsumoto et al., 2019; Mckenna et al., 2020; Okreglak and Walter, 2014; Qin et al., 2020). I introduced the knockouts of the above-mentioned targeting and quality control components

into mNG-TA and mNG-^{5}TMD^{5} libraries using synthetic genetic array (SGA) methodology (Baryshnikova et al., 2010; Tong et al., 2001) and assessed protein localization and abundance (Figure 23). ER and mitochondrial localization was validated by colocalization with the ER marker Sec61-mSc-I and the mitochondrial marker Cox4-mSc-I.

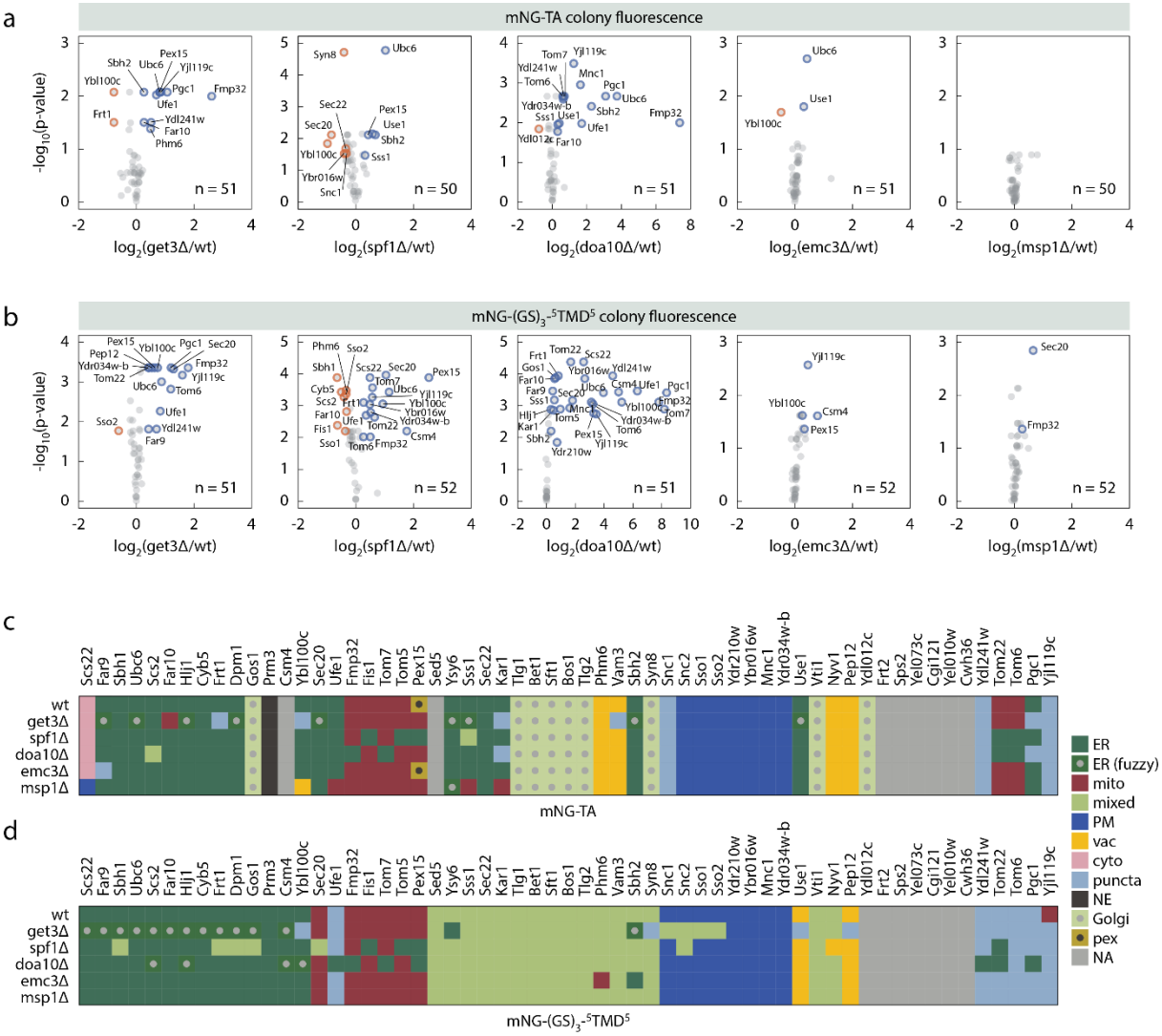

Figure 23. mNG-TMD fusions behave as TA proteins in the cell.

a , b - Log2 fold change of mNG fluorescence of N-terminally mNG-tagged TA proteins (a) or mNG-^{5}TMD^{5} fusions (b) expressed under a *GPD* promoter within *GET3*, *SPF1*, *DOA10*, *EMC3* or *MSP1* knockout background compared against wild type. Colony fluorescence measurements were taken from colonies grown on agar in 3 technical replicates and corrected for plate and spatial effects. Highlighted proteins show significant increase (blue) or decrease (red) in abundance ($p<0.05$). n – number of mNG-TA proteins or mNG-TMD variants analyzed.

c ,d - Localization of N-terminally mNG-tagged TA proteins (c) or mNG-^{5}TMD^{5} fusions (d) expressed under a *GPD* promoter in a wild type background or within *GET3*, *SPF1*, *DOA10*, *EMC3* or *MSP1* knockout background. ER, endoplasmic reticulum; ER (fuzzy), ER with disrupted shape; mito, mitochondria; mixed, several compartments; PM, plasma membrane; vac, vacuole; cyto, cytosol; NE, nuclear envelope; pex, peroxisome; NA, protein variants, which were not included in the experiment. For the variants without long C terminal extension, the mNG-^{5}TMD^{1} variants were used.

Comparison of *emc3Δ* and *get3Δ* mutants showed that the GET pathway is more important for TA protein targeting in yeast as 38% of mNG-TA proteins changed localization (Figure 23-c) and abundance (Figure 23-a) upon *GET3* deletion and only 6% upon *EMC3* deletion. The change in protein abundance, in this case, indicates that the insertion in the ER might be impaired and non-inserted proteins end up in a different compartment, form

aggregates or are degraded. A similar number of proteins changed localization (Figure 23-c,d) and abundance (Figure 23-a,b) in the full-length mNG-TA library (38%) and the mNG-^{5}TMD^{5} library (50%) upon *GET3* deletion with 48% of mNG-TA and corresponding mNG-^{5}TMD^{5} variants being affected in the same way, which shows that the recognition by the targeting factors is largely TMD specific. For instance, mNG-TA and mNG-^{5}TMD^{5} variants for Far9, Ubc6, Hlj1, Dpm1 and Sbh2 showed defects in the ER shape upon *GET3* deletion. mNG-Ybl100c and mNG-^{5}TMD^{5}(Ybl100c) variant showed change in localization from ER to puncta upon *GET3* deletion.

Knocking out different E3 ligases involved in the turnover of membrane proteins showed that Doa10 is very important for TA protein turnover (34% of mNG-TA and 54% of mNG-^{5}TMD^{5} fusion proteins changed localization and abundance upon *DOA10* deletion), which correlates with previously published data (Figure 23-b,d) (Dederer et al., 2019). Upon *DOA10* deletion, mitochondrial proteins Fmp32, Tom6, Tom7, Tom22 and mNG-^{5}TMD^{5} fusions with mitochondrial localization mNG-^{5}TMD^{5}(Fmp32), mNG-^{5}TMD^{5}(Tom7) and mNG-^{5}TMD^{5}(Pex15) changed the localization to the ER (Figure 24-a,b), which indicates that mislocalized mitochondrial proteins might be recognized and eliminated from the ER membrane with the help of Doa10.

Knocking out the protein dislocases Spf1 and Msp1 also confirmed the involvement of these regulators in TA protein homeostasis. 24% of mNG-TA and 46% of mNG-^{5}TMD^{5} fusion proteins changed localization and abundance upon *SPF1* deletion. Mitochondrial proteins Fis1, Tom5, Tom6, Tom22 and mNG-^{5}TMD^{5} fusions with mitochondrial localization mNG-^{5}TMD^{5}(Fis1), mNG-^{5}TMD^{5}(Tom5), mNG-^{5}TMD^{5}(Sec20) and mNG-^{5}TMD^{5}(Pex15) changed the localization to the ER upon *SPF1* deletion (Figure 24-a,b), which correlates with recently published data about Spf1 involvement in the removal of mislocalized mitochondrial proteins from the ER membrane (Mckenna et al., 2020). Only Sss1, Pex15 and Fmp32 accumulated at the mitochondria upon *MSP1* deletion in the mNG-TA screen (Figure 24-c). The presence of previously reported peroxisomal protein Pex15 (Li et al., 2019; Okreglak and Walter, 2014) among the hits confirms the involvement of the Msp1 in the removal of mislocalized peroxisomal proteins from the mitochondrial membrane. mNG-^{5}TMD^{5}(Fmp32) and mNG-^{5}TMD^{5}(Sec20) variants with mitochondrial localization showed significant increase in abundance upon *MSP1* deletion in mNG-^{5}TMD^{5} screen. This is in line with the previously reported accumulation of the full-length version of Fmp32 upon *MSP1* deletion (Dederer et al., 2019). These results indicate that Msp1 might control the levels of mitochondrial proteins in the outer mitochondrial membrane (OMM) in addition to the removal of mislocalized proteins.

In general, higher proportion of mNG-^{5}TMD^{5} fusions was affected upon QC factors deletion compared to mNG-TA proteins. Thus, 54% of mNG-^{5}TMD^{5} fusion proteins and 34% of mNG-TA proteins changed localization and abundance upon *DOA10* deletion and 46% of mNG-^{5}TMD^{5} fusion proteins and 24% of mNG-TA proteins changed localization and abundance upon *SPF1* deletion. These might be related to the fact that mNG-TMD fusions are artificial constructs, lacking any functional domains except the localization signal. The insertion of such

proteins into the membranes might prevent the interaction of other membrane proteins with the binding partners and QC system is able to recognize and eliminate mNG-^{5}TMD^{5} fusions. The analysis of the TMD hydrophobicity of Msp1, Spf1 and Doa10 substrate revealed that they tend to have the TMDs of low hydrophobicity compared to the variants, which were not affected by QC factors (Figure 24-d).

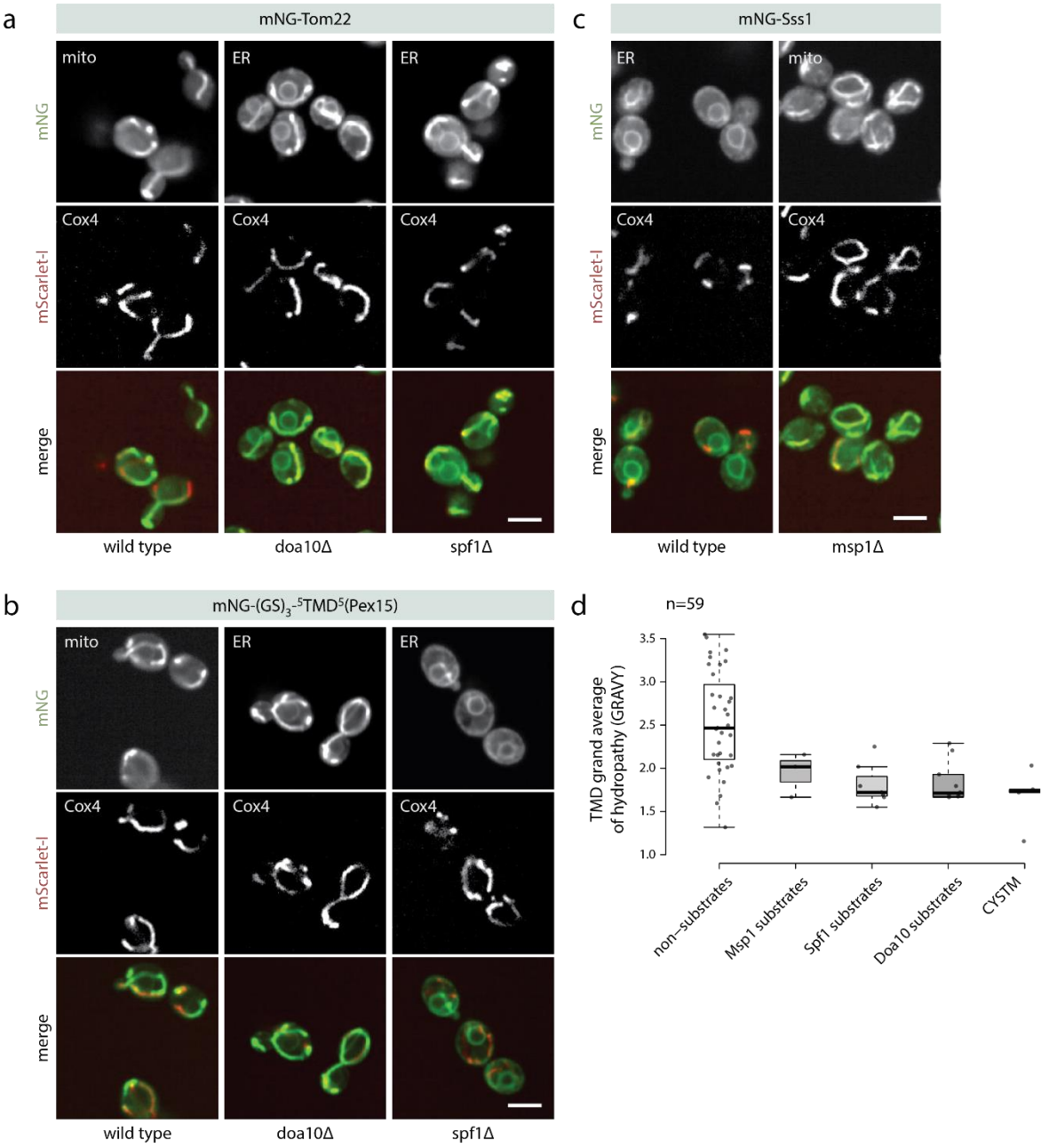

Figure 24. Change in localization for mNG-TA and mNG-^{5}TMD^{5} fusions upon the deletion of QC factors.
a – Mitochondrial protein mNG-Tom22 showed change in localization from mitochondria to ER in *doa10Δ* and *spf1Δ* mutants.
b – mNG-^{5}TMD^{5}(Pex15) variant showed change in localization from mitochondria to ER in *doa10Δ* and *spf1Δ* mutants.
c - mNG-Sss1 showed change in localization from ER to mitochondria in *msp1Δ* mutant. Localization class is indicated in the top left corner; Scale bars, 5 µm.
d – TMD GRAVY (grand average of hydropathy) of Msp1, Spf1 and Doa10 substrates compared to non substrates. CYSTM variants are separated from non substrates. Centerlines mark the medians, box limits indicate the 25th and 75th percentiles, and whiskers are the 95% confidence interval and any outliers are shown as circles. n – number of TMDs used in the analysis.

Overall, this genetic screen shows that mNG-^{5}TMD^{5} fusions behave as TA proteins in the cell and can be used as a model to study the sequence properties of the TMDs and the flanking regions. It is also a good demonstration of how synthetic protein variants generated using iPAL can be used for downstream applications.

5.3 MNG-TMD FUSIONS LOCALIZE AT DIFFERENT MEMBRANE-BOUND COMPARTMENTS

The results of the screen with different number of flanking residues (Figure 20) demonstrated that TMD alone can act as a localization signal, as 90% of TMD variants showed certain localization pattern and only 10% of the variants showed cytosolic localization. 68% of the mNG-TMD fusions anchored at the ER, mitochondria, PM, vacuole or several membrane compartments simultaneously (mixed category) and 22% of the mNG-TMD fusions showed puncta localization (Figure 25-d). Thus, Golgi, peroxisome and NE localization classes were not observed in the screen, even though the TMDs from the full-length proteins with these localization categories were included in the screen. This might indicate that extra localization signals are encoded somewhere else within the full-length sequence for these proteins. While for 30% of the variants (mostly ER category) the localization did not change upon the addition of the flanking residues, the rest of the variants showed different localization depending on the length of the flanking regions. This suggests that for a group of proteins specific targeting information is encoded in the TMD flanking regions.

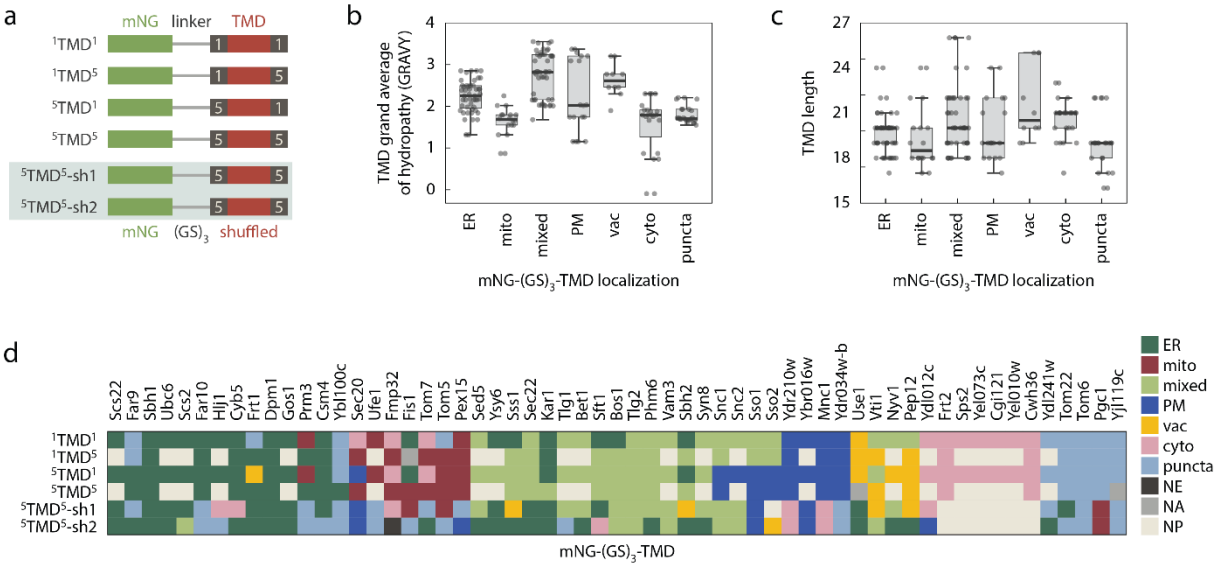

Figure 25. mNG-TMD fusions anchor at the different membrane compartments.
a - Library design. The TMD sequences from 59 yeast TA proteins, with the indicated number of native residues flanking the TMD (a single residue before and after the TMD (^{1}TMD^{1}), a single residue before and 5 residues after the TMD (^{1}TMD^{5}), 5 residues before and a single residue after the TMD (^{5}TMD^{1}), 5 residues before and 5 residues after the TMD (^{5}TMD^{5}) and with randomly shuffled TMD sequence (^{5}TMD^{5}-sh1 and ^{5}TMD^{5}-sh2)), were fused with the mNG.
b - TMD GRAVY (grand average of hydropathy) and localization of mNG-TMD fusions.
c – Length and localization of mNG-TMD fusions.
d – Localization of mNG-TMD fusions. ER, endoplasmic reticulum; mito, mitochondria; mixed, several compartments; PM, plasma membrane; vac, vacuole; cyto, cytosol; NE, nuclear envelope; NA, protein variants, which were not identified in the experiment; NP, protein variants without C-terminal extension, which were not included in the experiment.

To check the importance of the AA order within the TMD sequence I generated TMD variants with shuffled AA for the TMD region flanked by five AA from the native protein

sequence at the N- and C-terminal side (Figure 25-a highlighted in grey). To make sure that shuffled sequences still possess TMD properties, I used Phobius predictor (Käll et al., 2004) and checked if mNG-^{5}TMD^{5}-sh1 and mNG-^{5}TMD^{5}-sh2 variants are still predicted to have a TMD. 95% of mNG-^{5}TMD^{5}-sh1 and mNG-^{5}TMD^{5}-sh2 variants were predicted to have a TMD and the TMD start and end were similar to the predictions for the mNG-^{5}TMD^{5} variants. Interestingly, 73% of the mNG-TMD variants showing ER and mixed localization did not change the localization upon shuffling the AA within the TMD sequence (Figure 25-d), whereas 80% of the mitochondria and PM localized mNG-TMD variants changed the localization to a completely different compartment upon shuffling the AA within the TMD sequence (Figure 25-d). It is important to take into account that AA were scrambled randomly and different variants might have different degree of shuffling.

One of the important TMD features is hydrophobicity. I observed that TMDs of average hydrophobicity localize at the ER or throughout the endomembrane system (mixed category – variants localize at several compartments simultaneously), whereas mitochondrial TMDs have lower hydrophobicity and the PM group has 2 different trends – very high or low hydrophobicity (Figure 25-c). Most of the TMDs have a length of 18-20 AA but the PM group has 2 different trends – long (21-24 AA) or short (18-19 AA) TMDs (Figure 25-d). The puncta class has TMDs of low hydrophobicity. Such TMDs generally are not recognized by GET and, thus, might be more prone to aggregation.

Taken together, TMD alone can act as a localization signal, as 68% of mNG-TMD fusions demonstrated different membrane localization. The analysis of the TMD hydrophobicity and length suggests that these features of the TMD might be important for targeting specificity in combination with the presence of flanking residues.

5.4 IMPORTANCE OF POSITIVELY CHARGED FLANKING RESIDUES FOR LOCALIZATION

Comparison of the localization of the full-length proteins to the localization of the corresponding mNG-TMD fusions revealed several trends (Figure 26-a). About 30% of the TMDs do not depend on the flanking residues around hydrophobic stretch – even the shortest variants for these TMDs show the same localization as full-length proteins (Figure 26-e diagonal). About 40% of the variants show broad localization – they localize at several compartments simultaneously, including the expected compartment (Figure 26-f – mixed category). About 20% of the mNG-TMD fusions show completely different localization compared to the full-length protein in the absence of flanking residues (Figure 26-f, g, h). And for 9% of the dataset all TMD variants bring mNG to a completely different compartment compared to the full-length protein (Figure 26-a).

Comparison of the localization of mNG-^{1}TMD^{1} to mNG-^{5}TMD^{5} showed that for PM and mitochondria localized variants the presence of the flanking residues is important (Figure 26-b, mitochondrial and PM-localized variants highlighted in red). In particular, the residues at the C-terminus are important for mitochondrial localization (Figure 26-e,h). For example, the versions of the TMD from the mitochondrial protein Fis1 with C-terminal extension bring mNG-^{1}TMD^{5}(Fis1) and mNG-^{5}TMD^{5}(Fis1) to the mitochondria, whereas the versions of the

same TMD without flanking residues at the C-terminal part bring mNG-^{1}TMD^{1}(Fis1) and mNG-^{5}TMD^{1}(Fis1) to the ER (Figure 26-c).

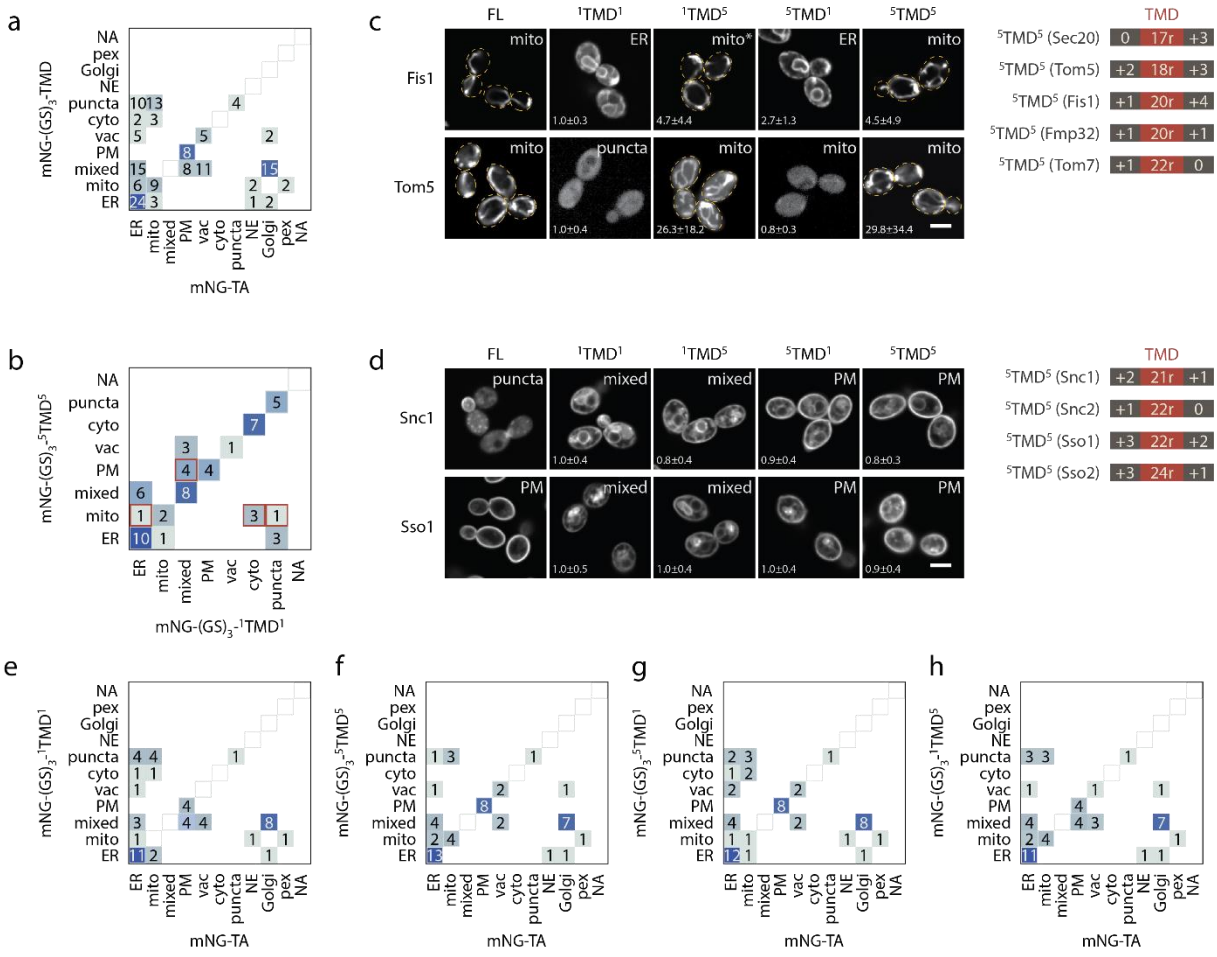

Figure 26. Importance of the flanking residues for localization.

a - Comparison of observed localization of all mNG-TMD fusions and full-length TA proteins localization.

b - Comparison of observed localization of mNG-^{1}TMD^{1} and mNG-^{5}TMD^{5}. For the variants without long C terminal extension the localization data for mNG-^{5}TMD^{1} was used. Mitochondrial and PM localized variants that changed the localization in the absence of flanking residues are highlighted in red.

c, d - Microscopy examples of change in localization in the absence of the flanking residues for the TMDs from mitochondrial TA proteins (c) and PM TA proteins (d). FL- full-length protein. Mean mNG intensity and SD normalized to the mNG intensity of the corresponding mNG-^{1}TMD^{1} variants are indicated in the bottom left corner; scale bar, 5 µm; localization class is indicated in the top right corner; the yellow dashed outlines indicate cell boundaries; schematic TMDs of the variants that changed localization in the absence of flanking residues with the indicated length and net charges of the flanking residues are shown on the right.

e, f, g, h - Comparison of observed localization of mNG-^{1}TMD^{1} (e), mNG-^{5}TMD^{5} (f), mNG-^{5}TMD^{1} (g), mNG-^{1}TMD^{5} (h) fusions and full-length TA proteins localization. For the variants without long C terminal extension the localization data for mNG-^{5}TMD^{1} was used. ER, endoplasmic reticulum; mito, mitochondria; mixed, several compartments; PM, plasma membrane; vac, vacuole; cyto, cytosol; NE, nuclear envelope; pex, peroxisome; NA, not identified.

Interestingly, for PM localization, the flanking residues at the N-terminus are important (Figure 26-e,g). For the TMD from PM localized protein Snc1 N-terminal extension seems to be important as mNG-^{5}TMD^{1}(Snc1) and mNG-^{5}TMD^{5}(Snc1) show higher PM signal than mNG-^{1}TMD^{5}(Snc1) and mNG-^{1}TMD^{1}(Snc1) (Figure 26-c,d). Figure 26-c,d shows the schematic TMDs of the mitochondrial and PM localized variants with the length of the TMD sequences and the net charge of the corresponding flanking residues. One common feature of these sequences

is the presence of positively charged AA within the flanking regions (Figure 26-c,d). Mitochondrial TMDs have shorter length and positive charge mainly at the C-terminus (Figure 26-c), and PM TMDs tend to be longer and have higher positive charge at the N-terminus (Figure 26-d).

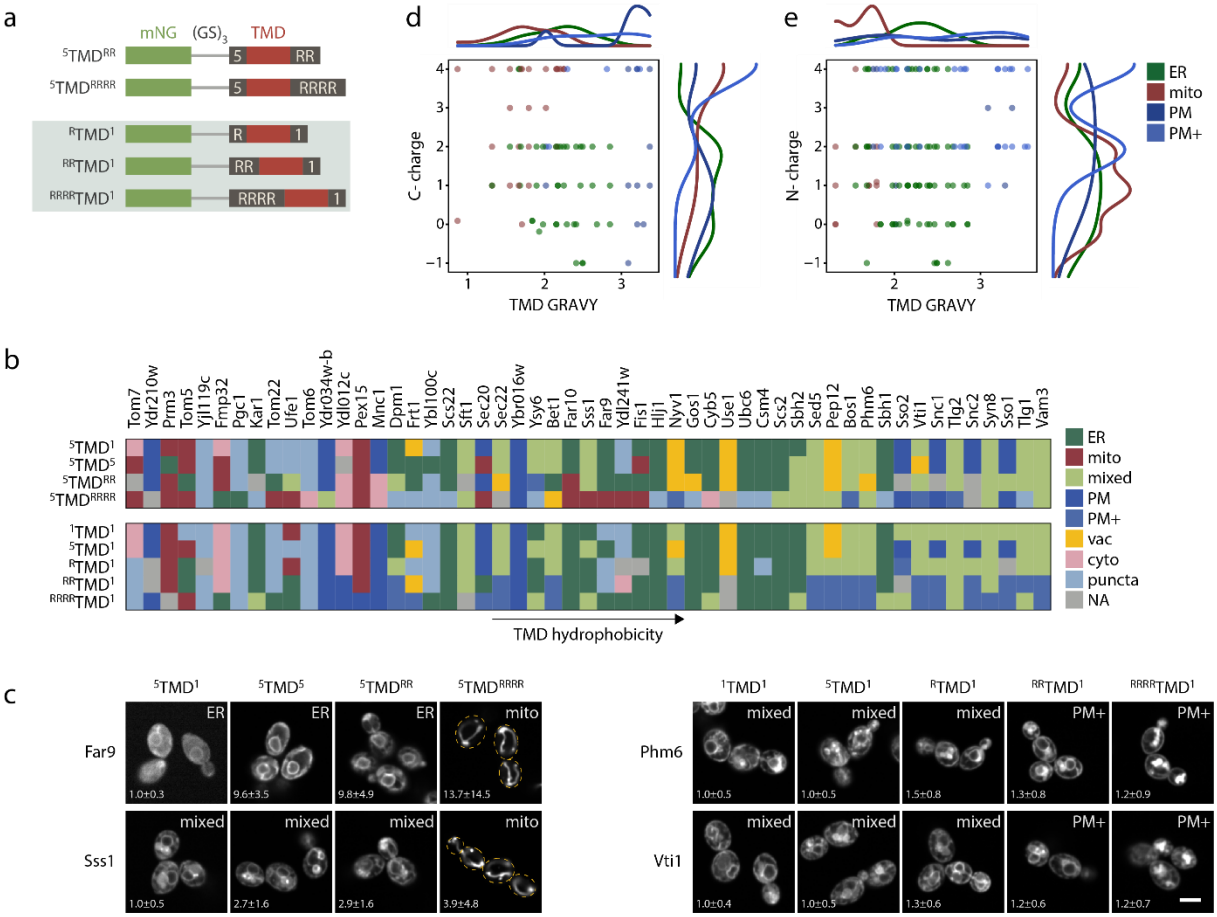

Figure 27. C- and N-terminal positively charged flanking residues are important for mitochondrial and PM localization.
a – Library design. 2 (^{5}TMD^{RR}) and 4 (^{5}TMD^{RRRR}) positively charged arginine residues were added to the C-terminal side of the TMD sequences from 59 yeast TA proteins, with 5 N-terminal flanking residues before the TMD (^{5}TMD) and fused with the mNG. 1 (^{R}TMD^{1}), 2 (^{RR}TMD^{1}) and 4 (^{RRRR}TMD^{1}) positively charged arginine residues were added to the N-terminal side of the TMD sequences from 59 yeast TA proteins, with 1 C-terminal flanking residue after the TMD (TMD^{1}) and fused with the mNG.
b - localization of mNG-TMD fusions with C- and N-terminal positively charged flanking residues compared to the TMD variants with a single residue before and after the TMD (^{1}TMD^{1}), 5 native residues before and a single native residue after the TMD (^{5}TMD^{1}) and 5 native residues before and after the TMD (^{5}TMD^{5}). The variants are ordered by TMD hydrophobicity.
c – Microscopy examples of the variants, which changed the localization upon the addition of the charge to the C-terminus and N-terminus. Mean mNG intensity and SD normalized to the mNG intensity of the corresponding mNG-^{5}TMD^{1} and mNG-^{1}TMD^{1} variants are indicated in the bottom left corner. Localization class is indicated in the top right corner. The yellow dashed outlines indicate cell boundaries; scale bar, 5 µm.
d, e – Dot plots with density overheads showing the localization of mNG-TMD fusions and the correlation between TMD GRAVY (grand average of hydropathy) and charge of C- (d) or N-terminal (e) flanking residues. Top curves show the distribution of TMD GRAVY, the curves on the right show charge scores distribution.
ER, endoplasmic reticulum; mito, mitochondria; mixed, several compartments; PM, plasma membrane; PM+, PM, ER and vacuole; vac, vacuole; cyto, cytosol; NE, nuclear envelope; NA, protein variants, which were not identified in the experiment.

To investigate how the presence of positively charged flanking residues affects the localization I generated mNG-TMD variants with two (mNG-^{5}TMD^{RR}) and four (mNG-^{5}TMD^{RRRR}) arginine residues at the C-terminus (Figure 27-a), and mNG-TMD variants with one (mNG-^{R}TMD^{1}), two (mNG-^{RR}TMD^{1}) and four (mNG-^{RRRR}TMD^{1}) arginine residues at the N-terminus (Figure 27-a). I have chosen arginine over lysine residues for this modification to avoid creating

a ubiquitination site on my synthetic variants. Thus, together with mNG-^{1}TMD^{1}, mNG-^{1}TMD^{5}, mNG-^{5}TMD^{1} and mNG-^{5}TMD^{5} my dataset contained the variants without charged flanking residues, with medium positive charge of 1 (R) or 2 (RR), and with very strong positive charge 4 (RRRR).

17% of mNG-TMD fusions changed the localization to mitochondria upon the presence of high charge (RRRR) at the C-terminus (Figure 27-b). All these variants had TMDs of low hydrophobicity (Figure 27-d, hydrophobicity less than 2.25). For instance, mNG-^{5}TMD^{1}(Far9) and mNG-^{5}TMD^{5}(Far9) with ER localization changed the localization to mitochondria upon the addition of 4 arginine residues to the C-terminus (Figure 27-c). A similar change in localization was observed for mNG-^{5}TMD^{1}(Sss1) and mNG-^{5}TMD^{5}(Sss1) with mixed localization (Figure 27-c).

In contrast, 30% of mNG-TMD fusions changed the localization to the PM or PM, ER and vacuole (PM+) upon the addition of medium (RR) and high (RRRR) N-terminal positive charge. Initially, these PM+ localized variants were classified as a mixed class by the automated classifier. Another round of visual inspection revealed the differences between mixed variants containing ER, Golgi and vacuole and mixed variants containing PM, ER and vacuole. The latter class was reannotated as PM+. Genetic validation of this reannotation is described further below (5.6). 70% of PM and PM+ localized variants had TMDs of high hydrophobicity (Figure 27-f, hydrophobicity more than 2.7). For example, mNG-^{1}TMD^{1} and mNG-^{5}TMD^{1} for Phm6 and Vti1 with mixed localization changed the localization to PM+ upon the addition of 2 and 4 arginine residues to the C-terminus (Figure 27-c). These findings are in line with the observations from the previous experiment with different number of flanking residues. Unexpectedly, the addition of high charge (RRRR) to the C-terminus of 10 TMDs that previously localized at the ER resulted in puncta localization of these mNG-^{5}TMD^{RRRR} variants. 50% of these variants had the TMDs of medium hydrophobicity and 50% had the TMDs of low hydrophobicity. These puncta structures did not colocalize with Golgi and peroxisomal markers.

Overall, these results suggest that the combination of low hydrophobicity of the TMD and the presence of positively charged residues at the C-terminus ensure mitochondrial localization, whereas high hydrophobicity of the TMD and the presence of positively charged residues at the N-terminus ensure PM localization.

5.5 IMPORTANCE OF TMD LENGTH AND HYDROPHOBICITY FOR LOCALIZATION

When I took a closer look at the TMD's features of PM-localized proteins there were two groups with different properties: long TMDs with high hydrophobicity and asymmetric distribution of valine residues, and shorter TMDs with low hydrophobicity enriched with cysteine residues (Figure 24-c,d). The properties of the first group were previously described in a computational study and were shown to be specific for the PM TMDs of fungal proteins (Figure 4) (Sharpe et al., 2010).

To test how important are TMD hydrophobicity and length for PM localization *in vivo* I elongated all TMD sequences to the total length of 22, 24 or 26 AA by adding V and A to the C-terminal part of the TMD (Figure 28-a).

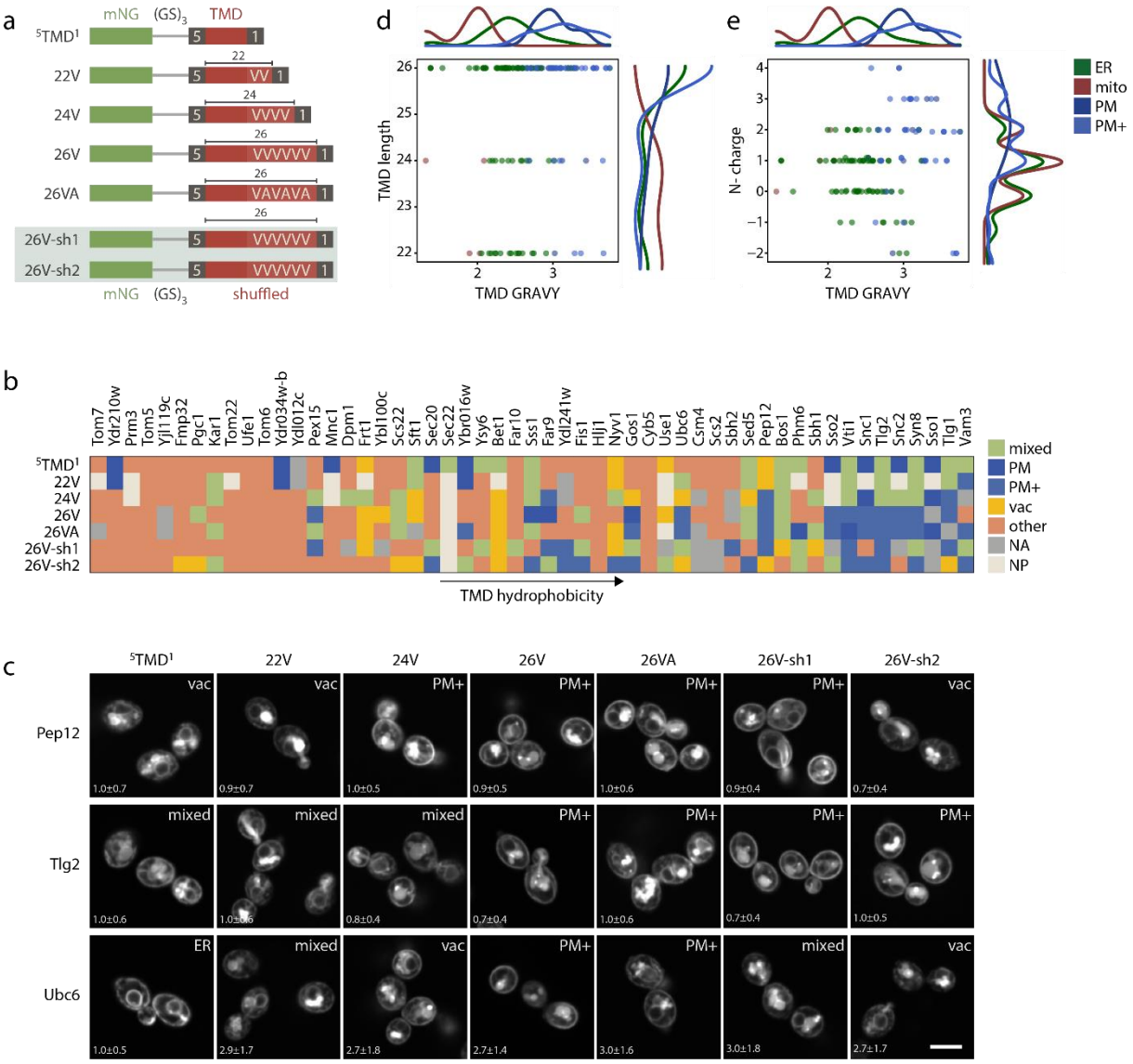

Figure 28. TMD elongation and increase of hydrophobicity lead to PM localization.
a – Library design. The TMD sequences from 59 yeast TA proteins with 5 native residues before and a single native residue after the TMD (^{5}TMD^1) were elongated by adding valine (V) and alanine (A) residues to the C-terminal part of the TMD to the total length of 22 (22V), 24 (24V) and 26 (26V and 26VA) AA. For the longest version of the TMD (26V), two combinations of randomly shuffled elongated TMD sequences (26V-sh1 and 26V-sh2) were created.
b - Localization of mNG-TMD fusions with the elongated TMDs compared to the TMD variants with 5 native residues before and a single native residue after the TMD (^{5}TMD^1). The variants are ordered by TMD hydrophobicity for the TMD variants with 5 native residues before and a single native residue after the TMD (^{5}TMD^1).
c – Microscopy examples of the variants, which changed the localization upon the TMD elongation. Mean mNG intensity and SD normalized to the mNG intensity of the corresponding mNG-^{5}TMD^1 variants are indicated in the bottom left corner. Localization class is indicated in the top right corner. Scale bar, 5 µm.
d, e - Dot plots with density overheads showing the localization of mNG-TMD fusions and the correlation between TMD GRAVY (grand average of hydropathy) and length (d) or charge of N-terminal (e) flanking residues. Top curves show the distribution of TMD GRAVY, the curves on the right show length (d) and charge scores (e) distribution.
ER, endoplasmic reticulum; mito, mitochondria; mixed, several compartments; PM, plasma membrane; PM+, PM and vacuole; vac, vacuole; cyto, cytosol; NE, nuclear envelope; NA, protein variants, which were not identified in the experiment; NP, TMD variants with the TMD length more than 22 residues, which were not included in the experiment.

Changing the length of the TMD by adding V residues to the C-terminal part resulted in the accumulation of mNG-TMD fusions at the PM and vacuole for 19% of TMD sequences (Figure 28-b,c). For example, mNG-^{5}TMD^{1}(Ubc6) variant localizes at the ER similar to the full-length protein but the extension of this TMD to 24 and 26 AA (24V, 26V and 26VA) changes the localization of the synthetic protein from ER to the PM+ (PM, ER and vacuole) (Figure 28-c). 70% of the variants that changed the localization to the PM and PM+ had TMDs of high hydrophobicity. Long TMDs of medium hydrophobicity showed ER localization (Figure 28-d). The analysis revealed that 80% of the variants that changed the localization to the PM and PM+ upon the TMD elongation had positively charged flanking residues at the N-terminus (Figure 28-e). This correlates with the previous screen, where the addition of positive charge to the N-terminus from the TMD resulted in the change of localization to the PM and PM+ for the highly hydrophobic TMDs. These results suggest that a combination of the hydrophobicity and presence of positive charges is required for targeting specificity and not a single property. To test this hypothesis, I picked 5 TMD variants (Sec20, Sbh1, Phm6, Scs2 and Bos1) with the longest extension (26V) that had very high hydrophobicity (more than 2.8) but no N-terminal charge and showed ER and mixed localization in the TMD length screen (Figure 29-a, b). Then I generated a set of variants of the same TMDs with the positive charge of 3 at the N-terminal side of the TMD (Figure 29-a). All the variants changed the localization to PM+ (Figure 29-b) confirming that the combination of high hydrophobicity and presence of positively charged flanking residues at the N-terminal side of the TMD ensures PM targeting specificity.

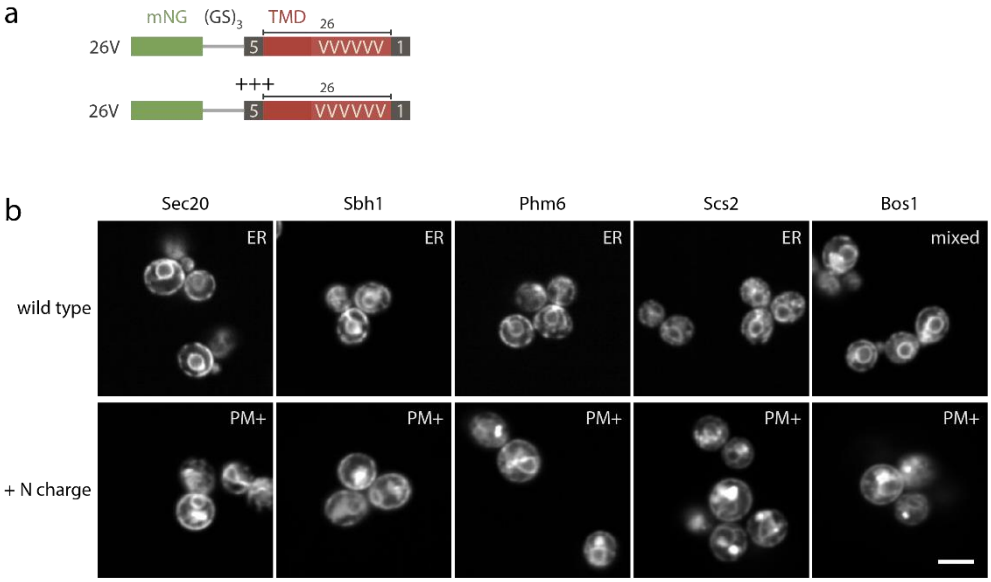

Figure 29. The presence of N-terminal positive charge is required for PM localization.
a – Oligo design. N-terminal charge of 3.0 was added to the 26V TMD variants.
b – Microscopy examples of change in localization from the ER and mixed to PM+ upon the addition of N-terminal charge. Localization class is indicated in the top right corner. Scale bar, 5 µm. ER, endoplasmic reticulum; mixed, several compartments; PM, plasma membrane; PM+, PM, ER and vacuole.

To check the importance of the asymmetric distribution of the valine residues and the importance of the AA order I generated the additional TMD variants with the shuffled AA order to balance the pool (Figure 28-a). 70% of the variants in this screen showed ER, mixed, PM and PM+ localization. The analysis of the fraction of each AA at the N- and C-terminal halves of the TMD, revealed an enrichment for V residues at the C-terminal half of the TMD for the PM and

PM+ localized variants. 83% of the variants had C-terminal fraction of V more than 0.3 (Figure 30-b), whereas for ER class only 51% of the variants had C-terminal fraction of V more than 0.3 (Figure 30-b). The mixed class shared the properties of ER and PM+ classes (Figure 30-f).

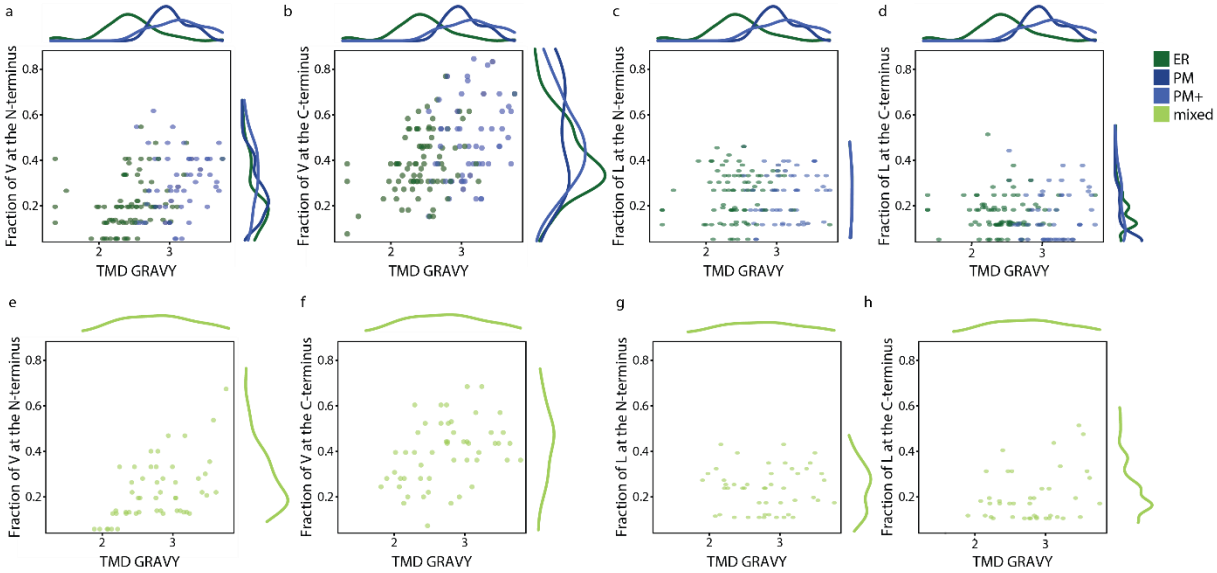

Figure 30. Importance of the AA composition for localization.

a, b, e, f – Dot plots with density overheads showing the localization of mNG-TMD fusions for ER, mixed, PM and PM+ classes and the correlation between TMD GRAVY (grand average of hydropathy) and fraction of V at the N-terminus (a, e) and fraction of V at the C-terminus (b, f). Top curves show the distribution of TMD GRAVY, the curves on the right show fraction of V at the N-terminus (a, e) distribution and fraction of V at the C-terminus (b, f) distribution.

c, d, g, h – Dot plots with density overheads showing the localization of mNG-TMD fusions for ER, mixed, PM and PM+ classes and the correlation between TMD GRAVY (grand average of hydropathy) and fraction of Lat the N-terminus (c, g) and fraction of L at the C-terminus (d, h). Top curves show the distribution of TMD GRAVY, the curves on the right show fraction of V at the N-terminus (c, g) distribution and fraction of V at the C-terminus (d, h) distribution. ER, endoplasmic reticulum; mixed, several compartments; PM, plasma membrane; PM+, PM, ER and vacuole.

In collaboration with , we have performed the analysis of different sequence properties on the combined dataset, which contained 816 TMD variants with the different number of flanking residues, variants with the extra N- and C-terminal charged flanking residues, variants with the alterations of the TMD length and shuffled TMD sequences. We have calculated TMD's hydrophobicity, AA composition, asymmetry in V and L residues, the charge of the flanking residues and residue volume within the TMD.

The analysis of the sequence properties of the combined dataset showed that low hydrophobicity and high C-terminal charge of the flanking residues are important for mitochondrial localization (Figure 31-a). Mitochondrial TMDs have short length (16-20 residues) (Figure 31-c). ER localization requires TMDs of medium hydrophobicity (Figure 31-a, b) and medium length (18-22 residues) (Figure 31-c). PM localized variants have long TMDs (22-26 residues) of high-hydrophobicity flanked with high N-terminal charge (Figure 31-b) or short CYSTM modules (18-20 residues) (Figure 31-f) of low hydrophobicity (Figure 31-d,e). The variants from mixed class localize at several compartments simultaneously, typically ER, Golgi and vacuole. The analysis of the sequence properties of this class revealed that this class shares the properties of ER and PM groups (Figure 31-d, e, f). Vacuolar class also shares the properties of ER and PM groups (Figure 31– g, h, i). The variants with cytosolic and puncta

localization have TMDs of low hydrophobicity without charged flanking residues (Figure 31–g, h, i).

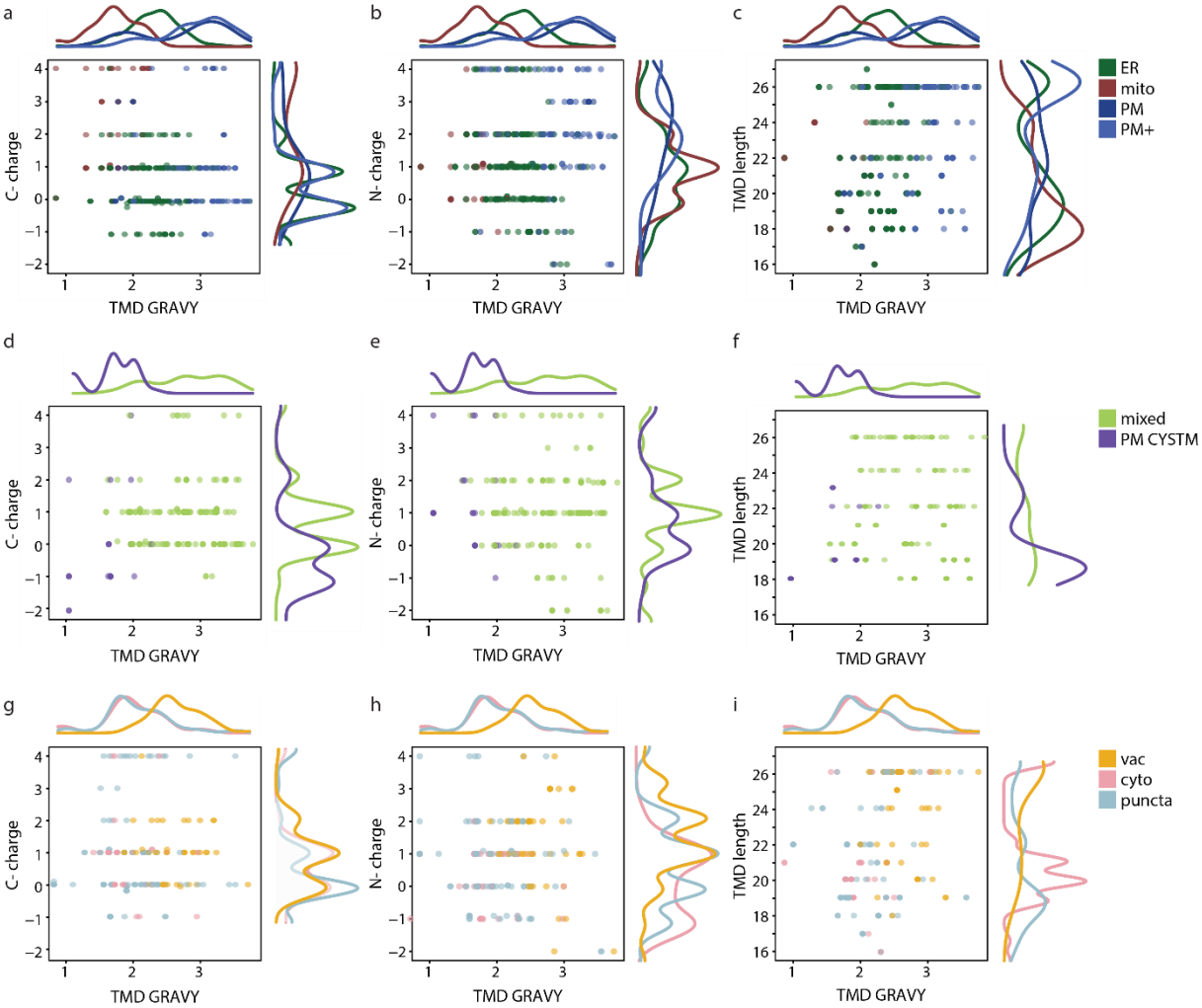

Figure 31. Biophysical properties of the TMDs important for protein localization.
a, b, c – Dot plots with density overheads showing the localization of mNG-TMD fusions for mitochondria, ER, PM and PM+ classes and the correlation between TMD GRAVY (grand average of hydropathy) and charge of C-terminal (a), N-terminal (b) flanking residues or TMD length (c). Top curves show the distribution of TMD GRAVY, the curves on the right show C-terminal (a) and N-terminal (b) charge scores distribution or TMD length distribution (c).
d, e, f – Dot plots with density overheads showing the localization of mNG-TMD fusions for mixed and PM CYSTM classes and the correlation between TMD GRAVY (grand average of hydropathy) and charge of C-terminal (d), N-terminal (e) flanking residues or TMD length (f). Top curves show the distribution of TMD GRAVY, the curves on the right show C-terminal (d) and N-terminal (e) charge scores distribution or TMD length distribution (f).
g, h, I – Dot plots with density overheads showing the localization of mNG-TMD fusions for vacuole, cytosol and puncta classes and the correlation between TMD GRAVY (grand average of hydropathy) and charge of C- (d), N-terminal (e) flanking residues or TMD length (f). Top curves show the distribution of TMD GRAVY, the curves on the right show C-terminal (d) and N-terminal (e) charge scores distribution or TMD length distribution (f). ER, endoplasmic reticulum; mito, mitochondria; mixed, several compartments; PM, plasma membrane; PM+, PM, ER and vacuole; CYSTM, variants with CYSTM module; vac, vacuole, cyto, cytosol.

Overall, these results suggest that a combination of TMD hydrophobicity and the charged flanking residues ensures correct targeting of mitochondrial and PM localized proteins, ER targeting requires the TMDs of medium hydrophobicity and a fraction of proteins

with the TMDs of medium and high hydrophobicity destined for residence at the vacuole and Golgi can migrate from the ER to these organelles by a vesicular trafficking route.

5.6 GENETIC VALIDATION OF PM LOCALIZATION

The images from the mixed category, which includes ER, Golgi and vacuole, and PM+ category, which includes PM, ER and vacuole, looked very similar (Figure 28-c). To make sure that these categories indeed represented localization at different compartments I performed additional genetic validation (Figure 32).

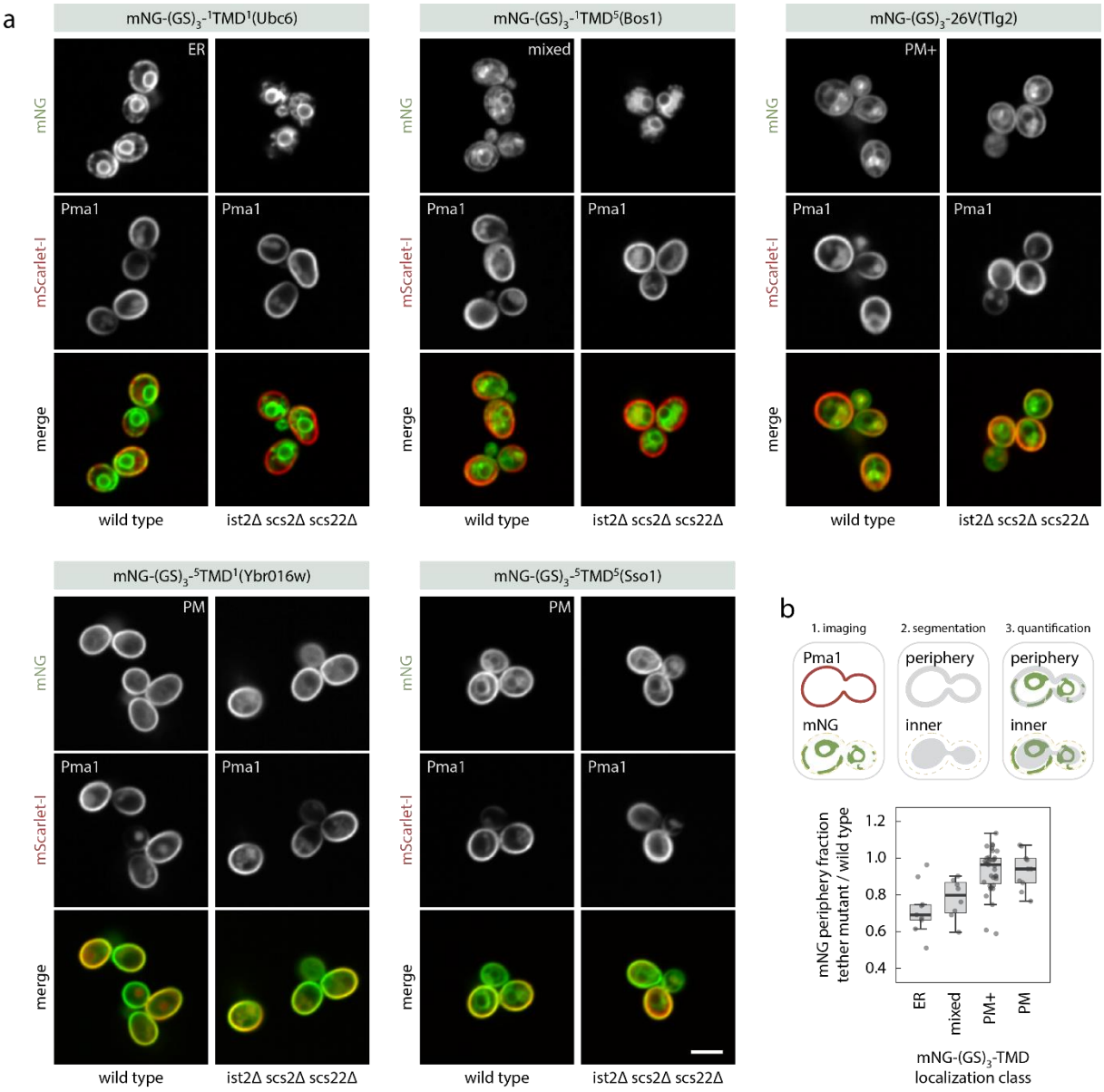

Figure 32. Genetic validation of PM+ localization class.

a – Change of the cortical ER shape in the triple tether mutant (*ist2Δ scs2Δ scs22Δ*) compared to the WT. ER and mixed strains show the disruption of the cortical ER in *ist2Δ scs2Δ scs22Δ* background, whereas PM and PM+ strains do not show any change of the phenotype. Localization class is indicated in the top left corner. Scale bar, 5 µm.

b – Quantification of the mNG intensity in the periphery fraction of the triple mutant in relation to the WT using Pma1-mSc-I as a segmentation marker. Cell periphery and inner part were segmented based on the signal of Pma1-mScarletI.

To differentiate between PM and cortical ER localization of mNG-TMD fusions, I created a strain yEI0164 with the knockout of three tethering factors *SCS2*, *SCS22* and *IST2*, and introduced the kanMX-GPDpr-mNG-TMD cassettes with 35 TMD variants with ER, PM, mixed and PM+ localization into this strain. It has been shown that cortical ER is disrupted in the triple deletion mutant due to the lack of PM-ER contact sites (Manford et al., 2012). Thus, only mNG-TMD fusions with the ER and mixed localization changed the phenotype in the mutant but not with PM and PM+ localization (Figure 32-a). yEI0164 also expresses the Pma1-mSc-I PM co-localization marker, which allowed the quantification of mNG signal intensity at the cell periphery. The quantification of the mNG ratio at the cell periphery and inside the cell in relation to WT shows a decrease in the mNG signal at the cell periphery for ER and mixed localized strains (Figure 32-b). This result proves that there are differences between mixed and PM+ localization categories and that our annotation of PM+ localization class is reliable.

5.7 DEEP MUTATIONAL SCANNING OF CYSTEINE-RICH TRANSMEMBRANE MODULE

The second group of PM-localized TMDs contained very short TMDs with low hydrophobicity enriched in cysteine residues. The enrichment in cysteine residues of PM-localized proteins was first identified by Thiago M. Venancio and L. Aravind using sensitive sequence profile analysis and referred to as CYSTM module (Figure 5) (Venancio and Aravind, 2010). Whether or not the CYSTM is a TMD is not confirmed experimentally. I decided to use iPAL scanning to check if cysteine residues are important for PM localization.

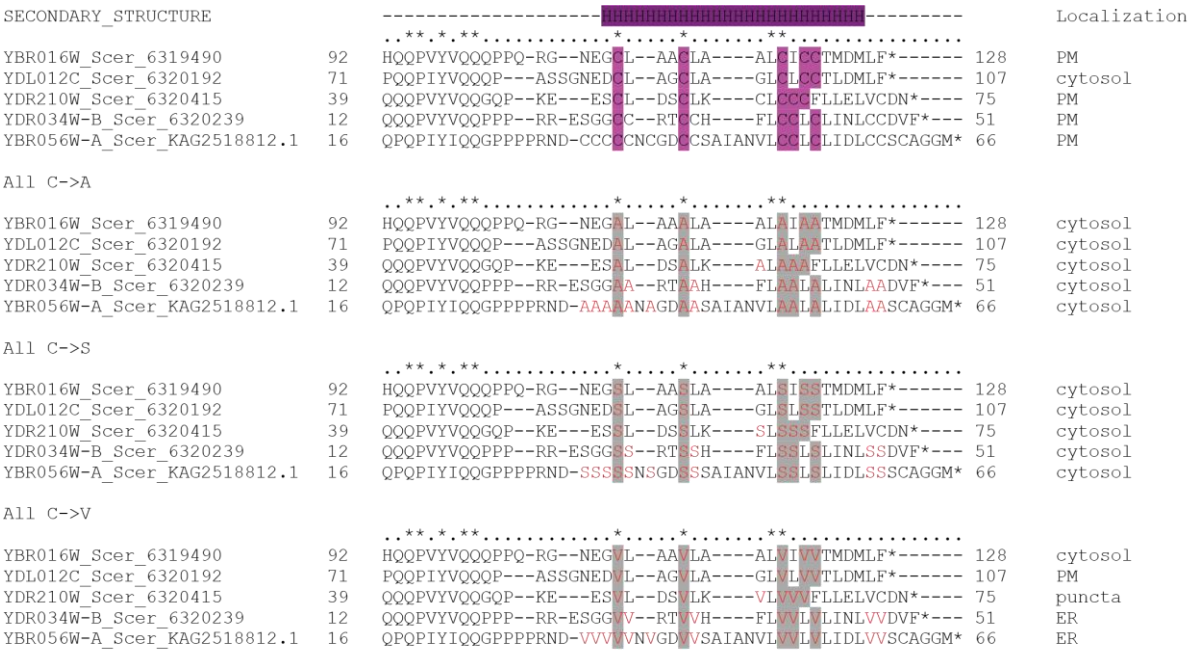

Figure 33. Mutations of C residues within CYSTM disrupt PM localization. Alignment of the C-terminal region of the CYSTM proteins and localization of mNG-TMD fusions. C residues in WT sequences are highlighted in purple. Mutated C positions are highlighted in grey. AA residues used for C substitutions are highlighted in red.

To test the importance of the cysteine motif for PM localization, I performed deep mutational scanning (DMS) of the TMDs from five CYSTM-containing proteins (Ybr016w, Ydr034w-b, Ybr056w-a, Ydr210w, Ydl012c). I constructed 379 sequence variants, where each

AA residue within the TMD was replaced by V or L, every single C was replaced by S, and all C were replaced by S, V or A. Replacing all C simultaneously by S or A resulted in cytosolic localization for all mNG-TMD fusions (Figure 33). Replacing all C simultaneously by V resulted in the change of localization to the ER, puncta and cytosol (Figure 33). Shuffling the AA order within the CYSTM region resulted in the change of localization from PM to puncta or cytosol (Figure 25-d). DMS did not reveal the particular importance of any individual C or other individual AA within the module.

To investigate if pairs or triplets of cysteine are important for PM-localization, I generated a series of mutants of Ybr016w with all possible pairs, triplets and quadruplets of C replaced by S (Figure 34). I also generated the variants where C positions within the TMD were fixed while other AA residues were shuffled and the variants where the complete TMD sequence was shuffled randomly.

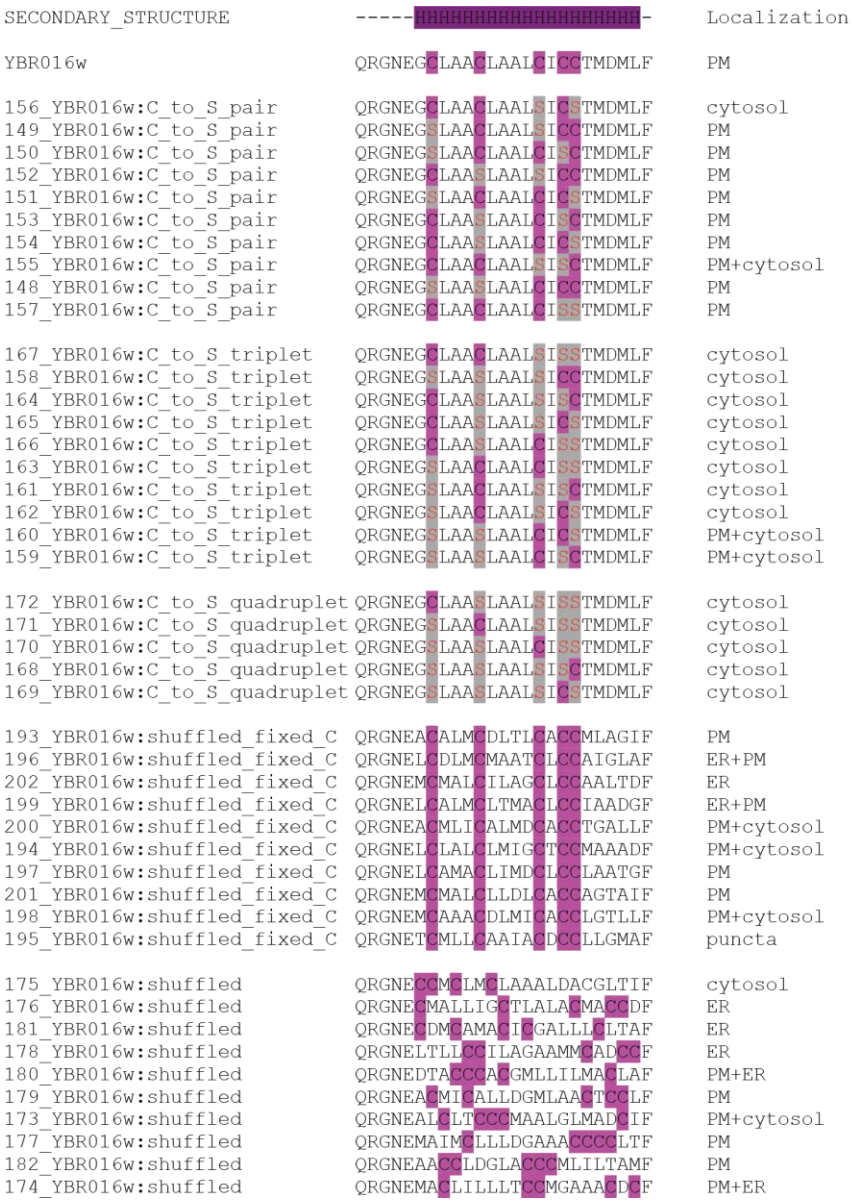

Figure 34. C residues at the C-terminal part of the CYSTM module are important for PM localization. DMS of Ybr016w. C residues are highlighted in purple. Mutated C positions are highlighted in grey. AA residues used for C substitutions are highlighted in red. The shuffling of the AA was performed within the helical domain indicated in dark purple.

The results show that mutating the pair of C to S in the C-terminal part of the CYSTM disrupts PM localization of the fusion protein in two combinations. A similar phenotype is present in the mutant with triple C mutated to S, with the third, fourth and fifth C residues in the module seem to have particular importance for PM localization. Mutating 4 out of 5 C residues to S completely disrupted PM localization. 70% of the shuffled variants with fixed C positions and 70% of randomly shuffled AA changed the localization, which suggest that some other AA contribute to the localization in addition to the cysteine residues at the C-terminal part of the module. This results also demonstrate how iPAL can be used for DMS experiments.

5.8 TAIL-ANCHORED PROTEINS CAN CARRY MORE THAN ONE LOCALIZATION SIGNAL

For the subset of mNG-TMD fusions that showed completely different localization compared to the full-length proteins (Figure 35-a), I analyzed the full-length protein sequences and checked the literature to find out if there are other localization signals present in the cytosolic part of the protein.

It has been demonstrated that full-length Pex15 mislocalizes to mitochondria upon the deletion of the last 30 AA at the C-terminus (Dederer et al., 2019; Okreglak and Walter, 2014). This implies that the peroxisomal targeting signal for this protein is located within these last 30 AA. The analysis of the full-length protein sequence of Sec20 revealed that it carries the ER-retention signal HDEL at the very C-terminus. Sequence analysis of the NE protein Prm3 with the NLS predictor (Nguyen Ba et al., 2009) showed that there is a non-canonical NLS at the positions 63-91. This suggests that some proteins carry several localization signals and targeting machinery might prefer some signals instead of the others.

The TMD of Sec20 has properties similar to the TMDs from mitochondrial proteins (low hydrophobicity, high positive C-terminal charge) and shows mitochondrial localization in fusion with mNG. However, full-length Sec20 localizes at the ER, probably, due to the presence of the ER-retention signal HDEL at the very C-terminus located 86 AA away from the TMD. I performed a small competition experiment to find out what happens upon fusing mitochondrial localization signal with the HDEL (Figure 35-b). I generated the following mutants: mNG-^{5}TMD^{5}(Sec20), mNG-^{5}TMD^{5HDEL}(Sec20) and mNG-^{5}TMD^{11}(Sec20). Interestingly, mNG-^{5}TMD^{5HDEL}(Sec20) variant still showed mitochondrial localization, whereas the presence of a small spacer of 6 AA coming from the Sec20 protein sequence in the variant mNG-^{5}TMD^{11}(Sec20) lead to ER localization (Figure 35-c). To test the minimum distance between the mitochondrial and ER-retention signal, I generated three variants with GS linkers of different length: mNG-^{5}TMD$^{5(GS)x3HDEL}$(Sec20), mNG-^{5}TMD$^{5(GS)x2HDEL}$(Sec20) and mNG-^{5}TMD$^{5(GS)HDEL}$(Sec20). The result shows that the spacer of at least 4 AA is required for the HDEL signal to be preferred instead of mitochondrial TMD.

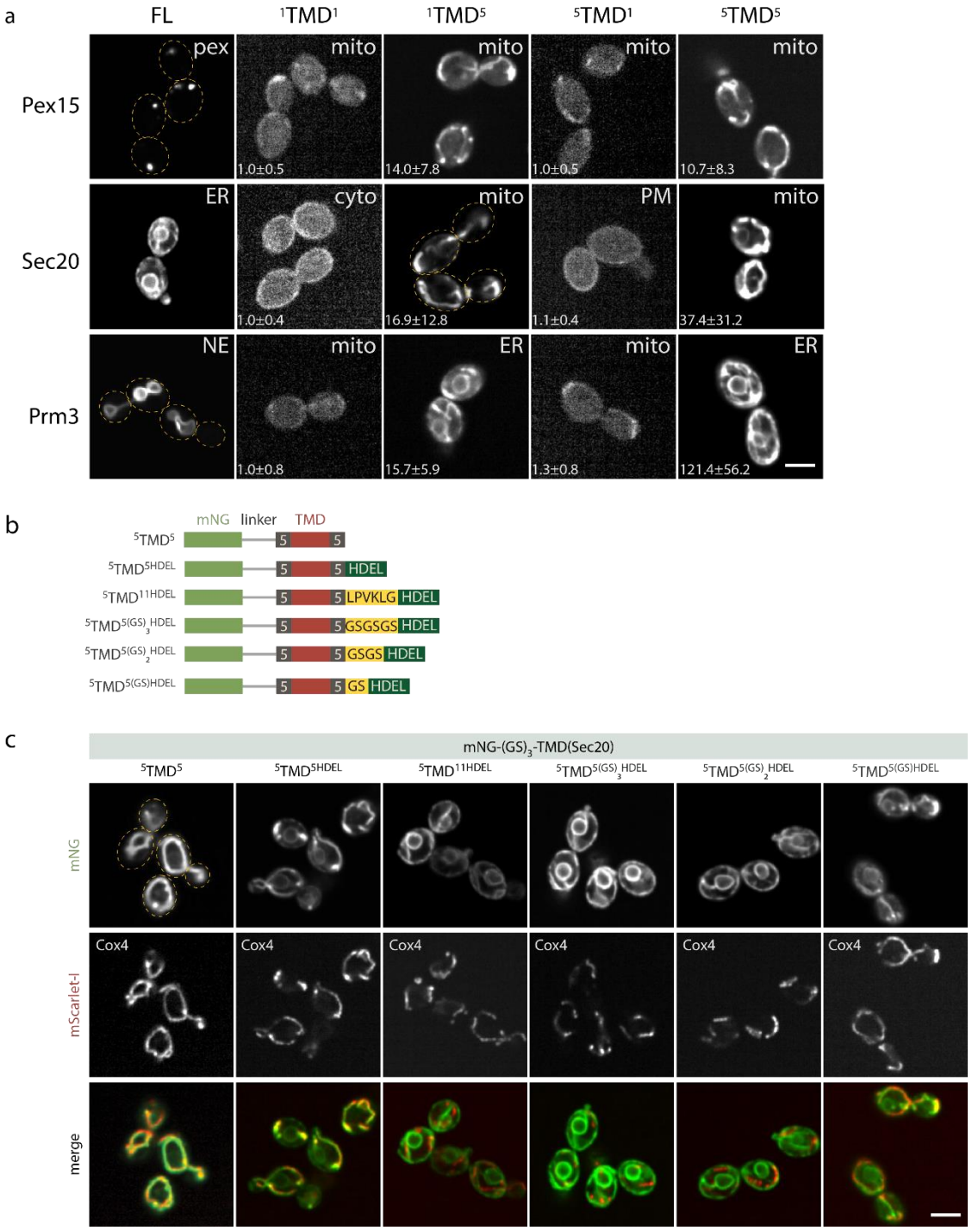

Figure 35. Targeting machinery can prefer one localization signal to the other.
a – Comparison of observed localization of mNG-TMD fusions and full-length (FL) TA proteins localization. Mean mNG intensity and SD normalized to the mNG intensity of the corresponding mNG-^{5}TMD^{1} variant are indicated in the bottom left corner. The yellow dashed outlines indicate cell boundaries; localization class is indicated in the top right corner; scale bar, 5 µm.
b – Design of the competition assay for mitochondrial TMD and ER retention signal. HDEL sequence was fused with ^{5}TMD^{5}(Sec20) directly or separated with 6 native residues from the native Sec20 sequence, (GS)x3, (GS)x2 or GS linker.
c- localization of mNG-TMD(Sec20) variants fused with HDEL and different linkers. Scale bar, 5 µm. ER, endoplasmic reticulum; mito, mitochondria; mixed, several compartments; PM, plasma membrane; cyto, cytosol; NE, nuclear envelope; pex, peroxisome.

These observations show that some tail-anchored proteins carry two different types of localization signals and the targeting machinery can prefer one signal to the other, probably, based on the position of the localization signals within the protein sequence.

6 DISCUSSION

6.1 IPAL HAS A WIDE RANGE OF APPLICATIONS

The initial aim of my PhD project was to expand the high-content screening toolbox by developing iPAL scanning, a deep mutational scanning approach to dissect protein localization signals. iPAL combines the high performance of pooled experiments with the advantages of downstream applications of arrayed libraries. In iPAL, pooled libraries of protein variants with alterations in a potential localization signal are converted into verified arrayed libraries using *in vivo* barcoding and deep sequencing. Due to the presence of a fluorescent reporter, high-throughput fluorescence microscopy can be used to determine the localization of each variant.

Using a set of TMDs from TA proteins, I demonstrated the efficiency and reproducibility of iPAL (Figure 20-b, c, Figure 21-c, e). The main advantage of iPAL compared to the recently developed visual screening methods based on photoconversion (Hasle et al., 2020; Kanfer et al., 2021; Yan et al., 2021) is that iPAL is not limited by the number of phenotypic bins and thus can be used to dissect complex localization signals. An automated image classification pipeline can identify 5 localization categories (ER, mitochondria, PM, vacuole and mixed) with high accuracy (Figure 21-c), which makes image analysis straightforward and quick. It is possible to modify the pipeline and retrain the algorithm to recognize more classes if needed.

iPAL has very high efficiency – 96% of the starting TMD variants were identified in the screens, and 95% of the identified variants had a biological replicate. Microscopy analysis of the biological replicates showed that the biological replicates for TMD variant show a very high correlation in terms of localization and mNG signal intensity (Figure 21-d, e). To demonstrate a proof of principle of iPAL, I included a subset of the TMD variants separated from mNG with Myc-(GS)x3 linker or Myc-(GS)x3-GGPG linker (Figure 20-a). Oligo variants with Myc-(GS)x3 and Myc-(GS)x3-GGPG linkers showed very similar localization (Figure 20-b). This result shows that Myc-(GS)x3 linker is neutral and does not affect TMD properties, as well as that the iPAL approach is reproducible and accurate.

In this study, I have demonstrated two different ways of applying iPAL. It is possible to perform massive parallel studies of different localization signals (TMDs from different proteins or different species) as well as canonical DMS with single amino acid substitutions (DMS of the CYSTM module) with localization as a read-out. The current reporter is ideal for investigating C-terminal localization signals. However, the system can be modified for investigating N-terminal localization signals. Therefore, iPAL is applicable for the detailed studies of different localization signals described in the introduction (NLS, PTS, MTS, etc.).

Another great advantage of iPAL compared to other modern methods for studying visual phenotypes is that all generated protein variants can be directly used for downstream applications, e.g., co-localization experiments or genetic screens. I have performed microscopy validation of all the variants in my screens by simply crossing them with different donor strains carrying co-localization and segmentation markers. I have also performed a genetic screen to identify TA protein turnover factors by crossing mNG-TMD variants with the

knockouts of the components of protein targeting and quality control systems. Another potential application of iPAL is to generate protein variants with mutations in a catalytic activity domain/interaction domain and perform downstream assays to test protein functionality.

The main limitation of iPAL is that it can only be used in *S. cerevisiae*, which makes it not compatible with visual cell sorting and IACS. However, the high accuracy and efficiency of iPAL are definitely attractive for the yeast genetics field.

6.2 MNG-TMD FUSIONS CAN BE USED TO STUDY TMD PROPERTIES IMPORTANT FOR THE LOCALIZATION

To make sure that mNG-^{5}TMD^{5} fusions behave as mNG-TA proteins, I introduced the knockouts of the known protein targeting and QC components *GET3*, *EMC3*, *TUL1*, *HRD1*, *ASI1*, *DOA10*, *MSP1* and *SPF1* into the mNG-TA and mNG-^{5}TMD^{5} libraries using synthetic genetic array (SGA) methodology and assessed protein localization and abundance (Figure 23). I observed that mNG-^{5}TMD^{5} fusions are affected by the knockouts of the protein targeting and turnover components in a similar way as mNG-TA. These results show that mNG-^{5}TMD^{5} fusions are recognized as TA proteins in the cell and can be used as a model to study the sequence properties of the TMDs and the flanking regions.

Comparison of *emc3Δ* and *get3Δ* mutants showed that the GET pathway is more important for TA protein targeting in yeast as more mNG-TA and mNG-^{5}TMD^{5} fusion proteins changed the localization (Figure 23-c) and abundance (Figure 23-a) upon *GET3* deletion. This is in line with the study from Bai and colleagues, where TMT mass spectrometry experiment identified only one TA protein among potential EMC substrates (Bai et al., 2020). Interestingly, it has been demonstrated that in mammalian cells EMC is required for the targeting of TA proteins with the TMDs of moderate hydrophobicity and multiple TA substrates were identified (Guna et al., 2018; Tian et al., 2019). This suggests that the role of EMC is more important in mammalian cells.

Since only about half of mNG-TA and mNG-TMD fusion proteins changed the localization and abundance upon *GET3* deletion, there is a possibility that GET and EMC pathways are partially redundant. To test this hypothesis the assessment of protein localization and abundance in the double mutant background is required.

6.3 DOA10 AND SPF1 ARE IMPORTANT FOR MAINTAINING ER HOMEOSTASIS

Knocking out different E3 ligases involved in the turnover of membrane proteins showed that Doa10 is very important for TA protein turnover, which correlates with previously published data (Figure 23-b,d) (Dederer et al., 2019). Due to the properties of my reporter (strong GPD promoter and bright fluorescent tag - mNG) I could identify some potentially new Doa10 substrates (mitochondrial proteins Tom22, Tom7, Tom6, peroxisomal protein Pex15 and ER residents Hlj1, Ufe1, Ybl100c) (Figure 23-c,d, Figure 24) in addition to previously reported ones (mitochondrial protein Fmp32 and ER residents Scs2, Pgc1, YDL241W) (Dederer et al., 2019). These proteins accumulated at the ER upon *DOA10* deletion.

Knocking out protein dislocases Msp1 and Spf1 also confirmed the involvement of these regulators in TA protein homeostasis. I could identify some potentially new Spf1 substrates (mitochondrial proteins Tom5, Tom6, Tom22) (Figure 23-c,d, Figure 24) in addition to the previously reported Fis1 (Keskin et al., 2017; Krumpe et al., 2012). Deletion of Spf1 results in the accumulation of these proteins at the ER. mNG-^{5}TMD^{5} fusions with mitochondrial localization mNG-^{5}TMD^{5}(Fis1), mNG-^{5}TMD^{5}(Tom5), mNG-^{5}TMD^{5}(Tom22), mNG-^{5}TMD^{5}(Sec20) and mNG-^{5}TMD^{5}(Pex15) also changed the localization to ER upon *SPF1* deletion. It has been hypothesized that Spf1 directly interacts with a positively charged lumenal segment of mitochondrial proteins and then flips the TMD out of the membrane (Mckenna et al., 2020). However, not all mitochondrial proteins have positively charged flanking residues. For instance, mNG-Tom22, mNG-Tom6 and mNG-^{5}TMD^{5}(Tom22) proteins that changed localization upon *SPF1* deletion in my screens (Figure 23-c,d, Figure 24) do not have a C-terminal charge. It would be interesting to test if there are more Spf1 substrates that do not possess a C-terminal charge and if they share other common features that might be important for the recognition by Spf1.

Interestingly, I found several mitochondrial proteins (Fmp32, Tom6, Tom7, Tom22) and mNG-^{5}TMD^{5} fusions with mitochondrial localization (mNG-^{5}TMD^{5}(Fmp32), mNG-^{5}TMD^{5}(Tom7) and mNG-^{5}TMD^{5}(Pex15)), which accumulated at the ER upon *DOA10* deletion, and three of them are also substrates of Spf1 (Tom6, Tom22 and mNG-^{5}TMD^{5}(Pex15)) (Figure 23-c,d, Figure 24). I also observed that all Spf1 and Doa10 substrates tend to have TMDs of low hydrophobicity (Figure 24). TMDs of low hydrophobicity are generally not recognized by the SRP and GET targeting factors and, thus, might be more subjected to the QC. These findings open the question if Spf1 and Doa10 cooperate in the removal of mitochondrial proteins from the ER membrane or if they are involved in two independent pathways. Another question is what the fate of mislocalized proteins after Spf1-dependent extraction is – re-insertion into the correct compartment or degradation? It would be interesting to use a time-lapse microscopy-based approach to answer these questions as Matsumoto and colleagues did for studying the fate of Msp1 substrates (Matsumoto et al., 2022). By combining inducible gene expression systems and time-lapse microscopy, Matsumoto and colleagues showed that authentic ER-TA proteins mistargeted to mitochondria return to the ER via the Msp1-GET pathway.

The tight collaboration between protein targeting and protein quality control systems is necessary to maintain correct targeting and protein homeostasis. Recent publications about protein dislocases and the observations in my screens raise the possibility that the ER serves as a default destination for all membrane proteins, and that protein dislocases are part of the targeting pathway instead of quality control components (Hansen et al., 2018; Matsumoto et al., 2022; Mckenna et al., 2020). Further studies are required for a better understanding of these processes. It would be important to test the behavior of the Spf1 substrates in *spf1Δ/doa10Δ* background as well as to test how the localization of mitochondrial proteins is changing in *spf1Δ* background upon the induction of *GET3* and *MSP1*.

6.4 BIOPHYSICAL PROPERTIES OF THE TMDS ENSURE CORRECT TARGETING OF TA PROTEINS

It has been shown for a few individual proteins that the TMD is important for the localization and that mutations in the TMD can lead to degradation, aggregation or protein mislocalization (Figueiredo Costa et al., 2018; Keskin et al., 2017; Okreglak and Walter, 2014; Vitali et al., 2018). However, the features of the TMDs crucial for targeting specificity are not well defined. Therefore, I applied iPAL to systematically dissect the features of the TMDs of all 59 yeast TA proteins important for targeting fidelity and constructed the library, containing the TMD variants with the different amounts of flanking residues. 68% of mNG-TMD fusions had membrane localization, which confirmed that the TMD alone can act as a localization signal (Figure 25).

Comparison of the localization of the mNG-TMD fusions with different length of flanking residues revealed that about 30% of the TMDs do not depend on the flanking residues around the hydrophobic stretch for their localization. Additional comparison with the corresponding full-length proteins demonstrated that all these variants show the same localization as full-length proteins (Figure 26-a diagonal). This group might be considered as “strong” localization signals since even the shortest variant of the TMD is enough to bring the reporter to the expected compartment (mostly including the TMDs from ER residents). Some of the variants show broad localization – they localize at several compartments simultaneously, including the expected compartment (Figure 25-d – mixed category). The analysis of TMD hydrophobicity showed that mNG-TMD fusions that localize at the ER or throughout the endomembrane system have TMDs of average hydrophobicity. Interestingly, most of the ER and mixed localized mNG-TMD variants did not change the localization upon shuffling the AA within the TMD sequence, suggesting that for these variants the average hydrophobicity of the TMD might be important and not a certain sequence pattern (Figure 25-d).

Additional sequence analysis of several mNG-TMD fusions for which all TMD variants showed completely different localization compared to the full-length proteins (Figure 35-a) revealed the presence of other localization signals within the full-length protein sequence. This suggests that some proteins carry several localization signals and that the targeting machinery might prefer some signals over others. The competition experiment between the TMD of Sec20 with the properties of mitochondrial TMDs and ER-retention signal HDEL showed that the spacer of at least 4 AA is required for the HDEL signal to be preferred instead of the mitochondrial TMD (Figure 35-b, c). This might be connected to the protein structure – possibly, with the shorter spacer HDEL signal is not exposed far enough from the α-helix and cannot be recognized by the HDEL receptor.

About 20% of the mNG-TMD fusions show completely different localization compared to the full-length protein in the absence of flanking residues (Figure 26-b, e mitochondrial and PM-localized variants highlighted in red). One common feature of these variants is the presence of positively charged AA within the flanking regions (Figure 26-c, d). In particular, the residues at the C-terminus are important for mitochondrial localization (Figure 26-e,f), which correlates with the previously published data for Fis1 protein – removing or mutating R and K

within Fis1 C-terminus led to the anchoring of Fis1 at ER (Rao et al., 2016). Interestingly, for PM localization, the positively charged residues are enriched at the N-terminus (Figure 26-c), which might be related to the topology of the protein and membrane properties and can potentially contribute to the targeting (Figure 26-g, h). The experiment with different number of flanking residues shows that the context around the TMD is important for the correct localization of mitochondrial and PM TA proteins (Figure 26). However, most of the mitochondria and PM localized mNG-TMD variants changed localization upon shuffling the AA within the TMD sequence, suggesting the presence of a certain sequence motif within the TMD that is responsible for localization of these variants in addition to the flanking residues (Figure 25-d).

Additional screens with different number of positively charged flanking residues at the C- and N-terminus of the TMD showed that mNG-TMD fusions with TMDs of low hydrophobicity changed the localization to mitochondria upon the presence of high charge (RRRR) at the C-terminus (Figure 27). This is in line with the previous study from Costello and collegues, where they demonstrated that TMDs of low hydrophobicity and a C-terminal extension with a net positive charge are targeted to mitochondria and peroxisomes (Costello et al., 2017). This result shows that a combination of biophysical properties is important for localization and not just a single feature. Unexpectedly, the addition of a high positive charge (RRRR) to the C-terminus of several TMDs that previously localized at the ER brought these mNG-^{5}TMD$^{5\ RRRR}$ variants to puncta, which did not colocalize with Golgi and peroxisomal markers.

In contrast, mNG-TMD fusions with TMDs of high hydrophobicity changed the localization to the PM upon the addition of a high N-terminal charge (RRRR). These findings are in line with the observations from the previous screen with the different number of flanking residues where long TMDs with high hydrophobicity and positive N-terminal charge showed PM localization in the fusion with mNG. The sequence analysis of the TMDs from single-span transmembrane proteins also demonstrated the enrichment in positively charged flanking residues at the cytosolic site and high hydrophobicity of the TMD for PM-localized transmembrane proteins (Sharpe et al., 2010). Changing the length of the TMD by adding V residues to the C-terminal part resulted in the accumulation of mNG-TMD fusions at the PM and vacuole for the variants with the TMDs of high hydrophobicity (Figure 28-b,c). This result indicates that longer length and high hydrophobicity might be required for the stable insertion of the TMD into the thick PM (Mitra et al., 2004; Sharpe et al., 2010).

The analysis of the sequence properties of the combined dataset with all generated variants (done by) confirmed that there is a correlation between hydrophobicity and positively-charged flanking residues important for the localization (Figure 31). Additionally, we have also checked AA composition, asymmetry in V and L residues and residue volume within the TMD, but this analysis did not reveal any additional features contributing for the localization. Overall, three localization classes can be distinguished based on the differences in hydrophobicity and positively charged flanking residues: mitochondrial localization requires low hydrophobicity of the TMD and a high positive charge at the C-terminal extension, ER

localization requires average hydrophobicity of the TMD and non-charged flanking residues, and PM localization requires high hydrophobicity of the TMD and high positive N-terminal charge or short CYSTM modules of low hydrophobicity. These features might be related to the differences in the membrane thickness and lipid composition for the various organelles (Krumpe et al., 2012; Mitra et al., 2004; Sharpe et al., 2010).

Hammond and colleagues show that plasmalemmal phosphatidylinositol 4-phosphate (PI4P) is involved in controlling protein recruitment and ion channel activity by electrostatic interactions (Fairn and Grinstein, 2012; Hammond et al., 2012). PI4P carries two or three negative charges at physiological pH and, thus, can recruit soluble polycationic proteins to the inner leaflet of the plasma membrane. PI4P can also bind to positively charged domains of the cytosolic aspect of transmembrane proteins during the fusion of vesicles. The enrichment of the TMDs flanked with positively charged residues in the PM fraction in my screens suggests that some TA proteins might be recruited to the PM by electrostatic interactions with PI4P.

The targeting route of TA proteins to mitochondria remains a mystery. The fact that the mitochondrial TMDs of low hydrophobicity flanked by positively charged residues are efficiently inserted into the mitochondrial membrane suggest the presence of an unknown targeting factor, which might recognize one of these properties or their combination. The systematic microscopic screen, in which the localization of a model OMM TA protein was visualized on the background of mutants in all yeast genes, revealed the change of localization from mitochondria to ER only upon *SPF1* deletion (Krumpe et al., 2012). It might worth trying to repeat this systematic microscopic screen in the double deletion background.

6.5 THE CYSTM MODULE IS IMPORTANT FOR PM LOCALIZATION

The detailed analysis of the TMDs from PM-localized variants revealed that there are two different trends – long TMDs of high hydrophobicity or short TMDs of low hydrophobicity (Figure 25-b, c). The latter class contains the variants with a short TMD enriched in cysteine residues (CYSTM) without charged flanking residues. In order to find whether cysteine residues are important for PM localization, I performed DMS of the CYSTM module from five CYSTM proteins. The results show that cysteine residues at the C-terminal part of the module are important for PM localization as mutating pairs of C to S in the C-terminal part of the CYSTM disrupts PM localization of the fusion protein (Figure 34). 70% of the shuffled variants with fixed C positions and 70% of randomly shuffled AA changed the localization, which suggest that some other AA contribute to the localization in addition to the cysteine residues at the C-terminal part of the module (Figure 34).

These results raise the possibility that the CYSTM module might have a different function. Very low hydrophobicity of the CYSTM module compared to the very high hydrophobicity of the other PM-localized transmembrane proteins contradicts the proposed model that the longer length and higher hydrophobicity are required for stable insertion of the TMD into the thick PM. Interestingly, replacing C by other non-hydrophobic AA within the low hydrophobic CYSTM region completely disrupted the localization of mNG-TMD fusions

(Figure 33). Replacing C by very hydrophobic V might have made it more similar to the canonical TMD sequence and resulted in different localizations but not the PM (Figure 33).

It is possible that cysteine residues are important for the oligomerization of CYSTM proteins (Mir and León, 2014; Xu et al., 2018) or for their interaction with TMDs from other PM proteins in the bilayer. However, it remains possible that CYSTM proteins are not integrated into the PM but are associated with the periphery. It would be important to conduct mass-spectrometry based proteomics to identify potential interactors of CYSTM proteins and perform fractionation assays to understand whether the CYSTM module can be inserted into the lipid bilayer.

6.6 CONCLUDING REMARKS

Taken together, in this study I described in detail the development of iPAL and the application of this novel visual screening approach to investigate localization determinants of TA proteins in *S. cerevisiae*. The resulting datasets of mNG-TMD variants localization and biophysical properties of the TMDs will enable future studies of targeting and QC mechanisms of transmembrane proteins. I anticipate that the iPAL approach will have a wide range of applications and help to understand the interplay between protein sequence and visual phenotypes.

7 MATERIAL AND METHODS

7.1 YEAST METHODS AND PLASMIDS

Yeast strains and plasmids used in this study are listed in Table S 1 and Table S 2, respectively. All yeast strains are in S288c background, derivatives of BY4741 and BY4742. Yeast genome manipulations (gene tagging and gene deletion) were performed using PCR targeting and lithium acetate transformation (Janke et al., 2004), unless stated otherwise. Experiments were performed at 30°C in synthetic complete (SC) medium with 2% (w/v) glucose as carbon source, unless stated otherwise.

Yeast tail-anchored proteins (Table S 3) were chromosomally tagged with mNeonGreen under the endogenous or the strong GPD promoters using the SWAp-Tag (SWAT) approach (Yofe et al., 2016). For that, N-SWAT acceptor strains were obtained from the N-SWAT library (Yofe et al., 2016). Using a pinning robot (Rotor, Singer Instruments) for sequential pinning and growth of yeast colonies on appropriate selective media following the synthetic genetic array (SGA) methodology (Baryshnikova et al., 2010; Tong et al., 2001), N-SWAT strains were mated with donor strains yEI0059 (carrying donor plasmid pEI098 for markerless tagging with mNeonGreen) or yEI0060 (carrying donor plasmid pEI099 for tagging with GPDpr-mNeonGreen), followed by sporulation, selection of haploids carrying both a donor plasmid and an N-SWAT acceptor locus, expression of the SceI endonuclease to promote tagging and selection of successful tagging events (Meurer et al., 2018; Weill et al., 2018).

Several reporter strains with different promoters for iPAL of TMDs in the context of tail-anchored proteins were created in the following way. The reporter strain yEI0011 with the GPD promoter was derived from strain JHY650 (Smith et al., 2017). First, the *SceI*$^{cut\ site}$*-FCY1-SceI*$^{cut\ site}$ cassette was removed from the *ybr209w*Δ locus. Subsequently, a GPDpr-mNeonGreen-GAGAGAGAGA-SceI$^{cut\ site}$-FCY1-SceI$^{cut\ site}$-T8term cassette for eventual expression of mNeonGreen-TMD fusions, controlled by the GPD promoter and the synthetic T8 terminator (Curran et al., 2015), was inserted into the *ybr209w*Δ locus between *kanMX* and *URA3pr* by PCR targeting using plasmids pEI087 as template. Successful transformants were selected on SC-ADE/URA agar plates (SC medium lacking adenine and uracil) (6.7 g/ l Bacto yeast nitrogen base without amino acids, 2 g/l amino acid dropout mix, 2% (w/v) glucose, 2% (w/v) agar) supplemented with 10 mg/L cytosine (BIC0713, Apollo Scientific) and validated by Sanger sequencing. Subsequently, GPDpr was replaced by a natNT2-TEF1pr cassette, amplified from plasmid pYM-N19, to create the strain yEI0024 with TEF1pr, and by a natNT2-ADH1pr cassette, amplified from plasmid pYM-N7, to create the strain yEI0025 with ADH1pr.

The final reporter strain yEI0038 with Myc-GSGSGS linker (Table S 1) for iPAL of TMDs in the context of tail-anchored proteins were derived from strain JHY650 (Smith et al., 2017). First, the *SceI*$^{cut\ site}$*-FCY1-SceI*$^{cut\ site}$ cassette was removed from the *ybr209w*Δ locus. Subsequently, a GPDpr-mNeonGreen-linker-SceI$^{cut\ site}$-FCY1-SceI$^{cut\ site}$-T8term cassette for eventual expression of mNeonGreen-TMD fusions, controlled by the GPD promoter and the

synthetic T8 terminator (Curran et al., 2015), was inserted into the *ybr209w*Δ locus between *kanMX* and *URA3pr* by PCR targeting using plasmid pEI091 (Myc-GSGSGS linker) as template.

The tether mutant strain yEI0164 (Table S 1) was used to differentiate between plasma membrane and cortical ER localization of mNeonGreen-TMD fusions. For that, kanMX-GPDpr-mNeonGreen-TMD cassettes were amplified by PCR with primers mNG-TMD-fw and mNG-TMD-rev from genomic DNA of individual iPAL library strains and integrated into the *his3*Δ1 locus of yEI0164.

7.2 CONSTRUCTION OF IPAL LIBRARIES

Oligonucleotide libraries encoding TMD variants (Table S 4) were designed for replacement by homologous recombination of the *SceI*$^{cut\ site}$*-FCY1-SceI*$^{cut\ site}$ cassette in the haploid reporter strain yEI0038 and were obtained as oligonucleotide pools (Twist Bioscience). Each pool (~70-100 ng of DNA) was amplified by PCR [Velocity DNA polymerase (Meridian Bioscience), 98°C for 2 min, (98°C for 30 s, 70°C for 30 s, 72°C for 5 s) × 20 cycles, 72°C for 5 min] using primers TMDpool-fw and TMDpool-rev (Table S 5) to elongate the homologous arms to 55 bp. 300 µL of each PCR product were purified using a PCR purification kit (Qiagen), transformed in the reporter strain yEI0038 and plated out on 5-FC plates (SC agar medium with 2% (w/v) glucose as carbon source and 1 g/L 5-fluorocytosine (5-FC, Apollo Scientific)). Transformations were scaled to obtain ~20x transformants for the number of TMD variants in a pool. After ~1-2 days of growth at 30°C, transformants were selected by size and circularity and arrayed on 5-FC plates in 384 format using a colony picking robot (PIXL, Singer Instruments; colony separation algorithm, radius filter 0.5-1.1, circularity filter 0.63-1.0). For some experiments oligos encoding TMD variants were ordered separately (Table S 6) and integrated into the reporter strain yEI0038 individually.

To identify the TMD variant in every transformant, arrayed transformants were mated in 1536 format, using a pinning robot (Rotor, Singer Instruments), with a barcoder library consisting of 1536 strains each carrying a different 26-nucleotide barcode in the *ybr209w*Δ locus (Smith et al., 2017). Diploids were pinned and grown on Raf/Gal plates (10 g/L Bacto yeast extract (ThermoFisher), 20 g/L Bacto peptone (ThermoFisher), 20 g/L Bacto agar (ThermoFisher), 2% (w/v) raffinose and 2% (w/v) galactose)) in two rounds to induce expression of the Cre recombinase and, through recombination, physically link the mNeonGreen-TMD and barcoder loci. Successful recombination events were selected in two rounds by growth on SC-URA plates (SC medium lacking uracil).

Diploid recombinants from each 1536-colony plate were pooled and total genomic DNA was extracted from ~150 million cells using a Gentra Puregene Yeast/Bact. Kit B (Qiagen) following the manufacturer's protocol. The genomic region of interest with a TMD linked to a barcode was amplified by PCR in two steps. A first PCR using 2 µg of genomic DNA as template and primers TMDamp_fw and TMDamp_rev (Table S 5) [NEBNextUltra DNA library prep kit (*New England Biolabs*), 98°C for 30 s, (98°C for 10 s, 60°C for 30 s, 72°C for 20 s) × 5 cycles, 72°C for 2 min] yielded a 747-809 bp amplicon. This product was gel-purified (QIAquick gel extraction kit, Qiagen) and used as a template for a second emulsion PCR with primers

P5_Scriptseq and P7-ix_Scriptseq (with a different 6-nucleotide index (ix) specific to each 1536-colony plate) (Table S 5) [98°C for 30 s, (98°C for 10 s, 58°C for 30 s, 72°C for 1 min) × 27 cycles, 72°C for 5 min]. Emulsion PCR was performed to minimize template switching (Williams et al., 2006). The resulting 827-889 bp amplicon was gel-purified and sequenced on a MiSeq platform using a MiSeq Micro flow cell and a 300-cycle kit (Illumina): 235 cycles for read1 (to determine the TMD sequence) and 80 cycles for read2 (to determine the barcode sequence and the plate index).

Analysis of sequencing data was performed in Python. Plate indexes and barcodes were matched to read2 sequences using Python's 'in' operator. Common before and after the TMD common to all variants were used to trim read1 using Python's string split method. Alternatively, cutadapt 3.1 package (Martin, 2011) was used to identify matching plate indexes, barcodes and trim read1. TMD sequences were matched to trimmed read 1 using Python's `in` operator. Reads were binned by plate-specific indexes and subsequently by the 26-nt barcodes specific to each position in the 1536-colony array. Only reads with perfect matches to the expected indexes and barcodes were considered. For each set of binned reads with more than 100 reads, the TMD sequence was then identified as the most commonly observed sequence (at least 65% frequency). Whenever possible, two independent clones were selected for each TMD sequence and the corresponding haploid strains in yEI0038 background were consolidated a new colony array using a colony picking robot (PIXL, Singer Instruments). Borders of the consolidated array were populated with yEI0038 as a dummy strain to minimize spatial growth effects within each plate.

7.3 GENETIC SCREENS

Using SGA (Baryshnikova et al., 2010; Tong et al., 2001), knockout alleles of *GET3*, *SPF1*, *EMC3*, *HRD1*, *ASI1*, *TUL1*, *MSP1*, *DOA10* or *URA3* as control were introduced into GPDpr-mNG-TA strains (MATalpha can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 hphMX-GPDpr-mNeonGreen-ORF) (Table S 3) by mating with strains yNS0010, yNS0011, yNS0012, UPS_hrd1, UPS_asi1, UPS_tul1, UPS_msp1, UPS_doa10, UPS_ura3 (Table S 1), followed by sporulation and selection of haploids carrying both a GPDpr-mNG-TA construct and a knockout allele on SD(MSG)-Leu/Arg/Lys + can/thi/G418/Hyg plates (1.7 g/ l Bacto yeast nitrogen base without amino acids and ammonium sulfate, 2 g/l amino acid dropout mix, 2% (w/v) glucose, 2% (w/v) agar, 1 g/l monosodium glutamic acid (MSG) (G1626, Sigma-Aldrich), 50 mg/l canavanine (C1625, Sigma-Aldrich), 50 mg/l thialysine (A2636, Sigma-Aldrich), 200 mg/l G418 (A291-25, Biochrom), 400 mg/l hygromycin B Gold (ant-hg-5, Invivogen)).

Similarly, GPDpr-mNG-$(GS)_3$-^{5}TMD^5 strains (MATalpha his3Δ leu2Δ lys2Δ ura3Δ ybr209w::KanMX-GPDpr-mNG-Myc-(GS)x3-^{5}TMD^5-T8term-URA3prom-5'URA3-Barcode-lox71-GalCre can1::MFa1pr-HIS3-MFalpha1pr-LEU2 fcy1::HphMX SAL1 CAT5(91M) MIP1(661T)) were crossed with strains yEI0112, yEI0113, yNS0015, yNS0016, yNS0017, yNS0019, yNS0020, yNS0021 and yNS0025 (Table S 1) and haploids carrying both a GPDpr-mNG-$(GS)_3$-^{5}TMD^5 construct and a knockout allele were selected on SD(MSG)-Leu/Arg/Lys + can/thi/G418/Nat plates (1.7 g/ l Bacto yeast nitrogen base without amino acids and

ammonium sulfate, 2 g/l amino acid dropout mix, 2% (w/v) glucose, 2% (w/v) agar, 1 g/l monosodium glutamic acid (MSG) (G1626, Sigma-Aldrich), 50 mg/l canavanine (C1625, Sigma-Aldrich), 50 mg/l thialysine (A2636, Sigma-Aldrich), 200 mg/l G418 (A291-25, Biochrom), 100 mg/l clonNAT (5.0, Werner BioAgents)).

The GPDpr-mNG-TA KO and GPDpr-mNG-^{5}TMD^{5} KO libraries were arranged in 1536-colony format using a pinning robot (RoToR, Singer Instruments). In each 4x4 quadrant, there are 3 technical replicates of the mNG-tagged strain or 3 technical replicates of the untagged control strain, each crossed with a KO strain and also with a WT strain. In each 4x4 quadrant, there are also 4 technical replicates of a reference strain. The reference strain has a strong mNG fluorescence signal and is spaced evenly across each plate and is present on all plates in the libraries. On each plate, dummy colonies are placed on the outer rows and columns to minimize the influence of nutrient access on colony size and fluorescence readout. Fluorescence intensities of the final colonies were measured after 24 h of growth using Infinite M1000 plate readers (TECAN) equipped with stackers for automated plate loading (TECAN) and custom temperature control chambers. Measurements in mNeonGreen (506/5 nm excitation, 524/5 nm emission, optimal detector gain) channel were performed at 400 Hz frequency of the flash lamp, with ten flashes averaged for each measurement.

Fluorescence measurement data were imported into R for data analysis. Data from dummy colonies were first removed and then mNeonGreen fluorescence intensities were normalized to the median of all reference strains on each plate to put all plates on the same scale. Fluorescence intensities were then corrected for spatial effects by normalizing to the median of the local reference strain in the 4x4 quadrant and corrected for cellular autofluorescence by subtracting the median of the untagged control strain. After all the normalizations, t-tests were performed to compare the knockout condition against the wild-type condition for each mNG-tagged strain and p-values were corrected using the Benjamini-Hochberg correction method. For plotting, mNeonGreen intensity values of technical replicates were then summarized by taking the mean and this final value was used to calculate the log2 fold change in intensity of knockout against wild type.

7.4 FLUORESCENCE MICROSCOPY

Haploid mNG-TMD libraries were mated with strain yEI0064 (Table S 1) expressing cytosolic tagBFP2 for cell segmentation and the ER marker Sec61-mScarlet-I. Strains with mNG-TMD variants localized to mitochondria or punctate structures were further examined by mating with strains yEI0063 (mitochondrial marker Cox4-mScarlet-I), yEI0072 (Golgi marker Anp1-mScarlet-I) and yEI0073 (peroxisomal marker Pex3-mScarlet-I) (Table S 1). To determine how mutations in protein targeting and turnover factors affect the localization of mNG-TMD fusions, mNG-$(GS)_3$-^{5}TMD^{5} and GPDpr-mNG-TA libraries (Table S 3) in wild type, *get3*Δ, *emc3*Δ, *msp1*Δ, *spf1*Δ and *doa10*Δ backgrounds were mated with strains mutated for the same targeting/turnover factors and expressing the mitochondrial marker Cox4-mScarlet-I or the ER marker Sec61-mScarlet-I and (yEI0063, yEI0064, yNS0086-yNS0095) (Table S 1). Diploid strains used for high-throughput fluorescence microscopy were selected on SD(MSG)-Ura+Hyg plates (1.7 g/ l Bacto yeast nitrogen base without amino acids and ammonium sulfate, 2 g/l amino

acid dropout mix, 2% (w/v) glucose, 2% (w/v) agar, 1 g/l monosodium glutamic acid (MSG) (G1626, Sigma-Aldrich), 400 mg/l hygromycin B Gold (ant-hg-5, Invivogen)) (Baryshnikova et al., 2010).

Diploid strains were grown to saturation in 96-well plates, diluted into fresh 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) and grown for 6 h to 0.5×10^7 cells/mL. 60 µL of culture per well were transferred to 384-well plates (PhenoPlate 384-well microplates, PerkinElmer) coated with 2 mg/mL concanavalin A (C2010, Sigma-Aldrich) solution in water (Khmelinskii and Knop, 2014). 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 10 fields of view (corresponding to an average of ~400 cells/well) were acquired with 2x2 pixel binning (corresponding to 300 nm pixel size) in the following order: mCherry (561 nm excitation, 570-630 nm emission filter), brightfield, tagBFP2 (405 nm excitation, 435-480 nm emission filter), mNeonGreen (488 nm excitation, 500-550 nm emission filter).

7.5 AUTOMATED IMAGE SEGMENTATION AND CLASSIFICATION

We developed a custom deep learning-based pipeline to classify the subcellular localization of mNeonGreen-TMD fusions. The pipeline segments cells in the BFP channel using a U-Net++ model (Zhou et al., 2018) and identifies the subcellular localization in the mNG channel using a Res-Net model (He et al., 2016) as follows:

– Input BFP channel images were normalized by dividing each pixel value with the median intensity of the corresponding pixel across all BFP images used to train the model. We clipped pixel values at the 99th intensity percentile, replacing higher values with that percentile, and linearly scaled the values between 0 and 255. To generate the ground truth data for training the cell segmentation model, ten randomly selected fields of view were segmented using ilastik (Berg et al., 2019) with human supervision. The segmentation masks were refined using the scikit-image package (Van Der Walt et al., 2014) to remove objects less than 100 pixels in size (morphology.remove_small_objects function) and to fill holes (morphology.remove_small_holes function).

– Next, we trained the U-Net++ model (Zhou et al., 2018), previously pre-trained to segment nuclei in mammalian cells (Ali et al., 2021), to segment yeast cells from 288 × 288 pixel patches of images in the BFP channel. Patches were randomly selected for training. The model was trained for up to 200 epochs with an initial learning rate of $3\text{x}10^{-4}$ and a mini-batch size of 8 to optimize binary cross-entropy loss using the Adam optimizer (Kingma and Ba, 2015). The learning rate was reduced by a factor of 10 once the validation loss was not improving for 10 consecutive epochs, and training was terminated if validation loss did not improve for 20 consecutive epochs.

– The trained U-Net++ model was applied to segment cells in sample BFP channel images. Sample images were pre-processed as training images, each image was split into 288 × 288 pixel patches, the segmentation masks were predicted by the model for each patch, and the masks were stitched back to a single image. The output segmentation masks were binarized and holes filled. Water-shedding was used to split the segmented objects into single cells. First, the seeds for water-shedding were obtained by considering the only pixels of the segmented objects that are 0.4 * the maximum distance from object pixel to the closest background pixel in the image. Water-shedding was applied starting from those seeds until all pixels of the segmented objects were covered. Objects smaller than 25 μm^2 or those with a circularity of less than 0.85, which likely correspond to out-of-focus buds, were filtered out.

– The subcellular localization model was trained to classify 100 × 100 pixel patches from the mNG channel centered on a cell. Cell masks were obtained from the segmentation model described above, and a patch was centered on the average of the extreme pixel coordinates in x and y directions for each cell segmentation mask. The pixel values in the input mNG images were linearly scaled between 0 to 255. We used the ResNet deep learning classification model (He et al., 2016) to distinguish between five localization categories: endoplasmic reticulum (ER), mitochondria (mito), plasma membrane (PM), vacuole (vac) and mixed. The mixed category is a complex phenotype with co-occurrence of mostly ER, vac and PM localizations in the same cell. To train the localization model, 170 and 30 manually annotated fields of view per localization category were used for training and validation, respectively. The model was trained for up to 200 epochs with an initial learning rate of $5x10^{-4}$ and a mini-batch size of 128 to optimize the binary cross-entropy loss using the Adam optimizer (Kingma and Ba, 2015). The learning rate was reduced by a factor of 10 when validation loss did not improve for 10 consecutive epochs, and training was terminated if validation loss did not improve for 20 consecutive epochs.

– The trained ResNet model was applied to sample mNG images with the same scaling and patching. To determine the protein localization category of an image, we averaged the localization probability of the patches in this image for each category and assigned the category with the highest average probability to the image. In addition, we quantified the average single-cell mNG signal for each strain. For that, we subtracted the median image intensity in the GFP channel, removed the 2% brightest and dimmest cells, and then calculated the average pixel intensity for each cell.

– The model architectures were built using the Keras framework with TensorFlow backend 2.3.0 (Abadi et al., 2016) and the experiments were run on Tesla V100-PCIE-GB Graphics Processing Units at The High-Performance Computing Center of the University of Tartu.

7.6 ANALYSIS OF TMD PROPERTIES

To minimize linker influence on TMDs properties I have tried to identify the most neutral linker computationally by testing how the different linker sequences between mNeonGreen and TMD might affect predictions of the TMD's position. 50 yeast TA proteins'

sequences, which TMDs can be identified by Phobius prediction tool (Käll et al., 2004), and corresponding mNG-linker-TMD synthetic fusions were analyzed using Phobius prediction tool. Difference in TMD's predicted start and end were compared for (GA)x5, (GS)x3 and GGPG linkers using R.

We analyzed the patterns in the physical and physiochemical property of the TMDs of the 59 peptides under study with their localization. Peptides were represented using five physical properties, namely fraction of Valine at N Terminal half, fraction of Valine at C Terminal half, Fraction of Leucine at N terminal half and Fraction of Leucine at C terminal half and Length of the TMD, and three physiochemical property - Hydrophobicity, N Terminal Charge, C Terminal Charge. Hydrophobicity of the TMD was calculated using the KyteDooLittle scale and the charges were calculated using the "Zimmerman" scale from the "peptides" package in R. The physical properties were calculated using string functions in R. Visualization was done using "ggplot2" library in R.

8 APPENDIX I

8.1 SUPPLEMENTARY

Table S 1. Yeast strains.

Strain	Background	Genotype	Source
BY4741	S288c	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	(Brachmann et al., 1998)
BY4742	S288c	MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0	(Brachmann et al., 1998)
Y8205	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2	(Tong and Boone, 2007)
yMaM639	Y8205	leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2	(Meurer et al., 2018)
JHY650	BY4742	ybr209wΔ::kanMX-URA3pr-URA3ΔC-barcode-SceI$^{cut\ site}$-FCY1- SceI$^{cut\ site}$-lox71-GALpr-Cre can1Δ::MFa1pr-HIS3-MFalpha1pr-LEU2 fcy1Δ::hphMX SAL1 CAT5(91M) MIP1(661T)	(Smith et al., 2017)
SHA349	BY4742	ybr209wΔ::NatMX-GalCre can1Δ::MFa1pr-HIS3-MFalpha1pr-LEU2	(Smith et al., 2017)
SHA345	BY4741	ybr209wΔ::GalCre-NatMX can1Δ::MFa1pr-HIS3-MFalpha1pr-LEU2	(Smith et al., 2017)
#2797	BY4742	ybr209wΔ::KanMX-URA3prom-5'URA3-Barcode-lox71-GalCre can1Δ::MFa1pr-HIS3-MFalpha1pr-LEU2	(Smith et al., 2017)
#2836	BY4742	ybr209wΔ::KanMX-URA3prom-5'URA3-Barcode-lox71-GalCre can1Δ::MFa1pr-HIS3-MFalpha1pr-LEU2 fcy1Δ::HphMX	(Smith et al., 2017)
#2849	BY4742	ybr209wΔ::KanMX-URA3prom-5'URA3-Barcode- SceI$^{cut\ site}$ -FCY1prom-FCY1- SceI$^{cut\ site}$ -lox71-GalCre can1Δ::MFa1pr-HIS3-MFalpha1pr-LEU2 fcy1Δ::HphMX	(Smith et al., 2017)
N-SWAT_TA	BY4741	L1-hphΔN-SceI$^{cut\ site}$-URA3-SpNOP1pr-sfGFP-L2::ORF	(Yofe et al., 2016)
yEI0011	JHY650	ybr209wΔ::kanMX-GPDpr-mNeonGreen-GAGAGAGAGA-SceI$^{cut\ site}$-FCY1-SceI$^{cut\ site}$-T8term-URA3pr-URA3ΔC-barcode-lox71-GALpr-Cre	this study
yEI0024	yEI0011	ybr209wΔ::kanMX-natNT2-TEF1pr-mNeonGreen-GAGAGAGAGA-SceI$^{cut\ site}$-FCY1-SceI$^{cut\ site}$-T8term-URA3pr-URA3ΔC-barcode-lox71-GALpr-Cre	this study
yEI0025	yEI0011	ybr209wΔ::kanMX-natNT2-ADH1pr-mNeonGreen-GAGAGAGAGA-SceI$^{cut\ site}$-FCY1-SceI$^{cut\ site}$-T8term-URA3pr-URA3ΔC-barcode-lox71-GALpr-Cre	this study
yEI0059	yMaM639	pEI098 (pSD-N1-mNeonGreen)	this study
yEI0060	yMaM639	pEI099 (pSD-N4-GPDpr-mNeonGreen)	this study
yEI0038	JHY650	ybr209wΔ::kanMX-GPDpr-mNeonGreen-Myc-GSGSGS-SceI$^{cut\ site}$-FCY1-SceI$^{cut\ site}$-T8term-URA3pr-URA3ΔC-barcode-lox71-GALpr-Cre	this study
yEI0063	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term COX4-mScarlet-I-KanMX	this study
yEI0064	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term SEC61-mScarlet-I-KanMX	this study
yEI0072	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term ANP1-mScarlet-I-KanMX	this study
yEI0073	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term PEX3-mScarlet-I-KanMX	this study
yNS0086	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term COX4-mScarlet-I-KanMX msp1Δ::natNT2	this study
yNS0087	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term COX4-mScarlet-I-KanMX get3Δ::natNT2	this study
yNS0088	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term COX4-mScarlet-I-KanMX spf1Δ::natNT2	this study
yNS0089	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term COX4-mScarlet-I-KanMX emc3Δ::natNT2	this study
yNS0090	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term COX4-mScarlet-I-KanMX doa10Δ::natNT2	this study
yNS0091	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term SEC61-mScarlet-I-KanMX msp1Δ::natNT2	this study
yNS0092	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term SEC61-mScarlet-I-KanMX get3Δ::natNT2	this study
yNS0093	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term SEC61-mScarlet-I-KanMX spf1Δ::natNT2	this study
yNS0094	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term SEC61-mScarlet-I-KanMX emc3Δ::natNT2	this study
yNS0095	BY4741	leu2Δ::URA3-GPDpr-tagBFP2-CYC1term SEC61-mScarlet-I-KanMX doa10Δ::natNT2	this study
yNS0010	BY4741	get3Δ::kanMX	this study
yNS0011	BY4741	spf1Δ::kanMX	this study
yNS0012	BY4741	emc3Δ::kanMX	this study
UPS_hrd1	BY4741	hrd1Δ::kanMX	(Kong et al., 2021a)
UPS_asi1	BY4741	asi1Δ::kanMX	(Kong et al., 2021a)
UPS_tul1	BY4741	tul1Δ::kanMX	(Kong et al., 2021a)
UPS_msp1	BY4741	msp1Δ::kanMX	(Kong et al., 2021a)

Strain	Background	Genotype	Source
UPS_doa10	BY4741	doa10Δ::kanMX	(Kong et al., 2021a)
UPS_ura3	BY4741	ura3Δ::kanMX	(Kong et al., 2021a)
yEI0112	BY4741	lyp1Δ0 hrd1Δ::natNT2	this study
yEI0113	BY4741	lyp1Δ0 msp1Δ::natNT2	this study
yNS0015	BY4741	lyp1Δ0 get3Δ::natMX	this study
yNS0016	BY4741	lyp1Δ0 spf1Δ::natMX	this study
yNS0017	BY4741	lyp1Δ0 emc3Δ::natMX	this study
yNS0019	BY4741	lyp1Δ0 tul1Δ::natMX	this study
yNS0020	BY4741	lyp1Δ0 doa10Δ::natMX	this study
yNS0021	BY4741	lyp1Δ0 asi1Δ::natMX	this study
yNS0025	BY4741	lyp1Δ0 his3Δ::natMX	this study
yEI0164	BY4741	scs2Δ::natNT2 scs22Δ::hphNT1 ist2Δ::CaURA3 PMA1-mScarlet-I-HIS3MX6	this study

Table S 2. Plasmids.

Plasmid	Description	Source
pSD-N1	donor template for markerless tagging using N-SWAT (L1-MCS-L2), kanMX4 plasmid selection	(Yofe et al., 2016)
pSD-N4	donor template for tagging using N-SWAT with reconstitution of the hphNT1 marker (hphΔC-MCS-L2), kanMX4 plasmid selection	(Yofe et al., 2016)
pEI098	pSD-N1-mNeonGreen (N-SWAT donor type I)	this study
pEI099	pSD-N4-GPDpr-mNeonGreen (N-SWAT donor type II)	this study
p415GPD	pBLUESCRIPT II shuttle vector, AmpR, CEN6 ARSH4 LEU2, GPDpr	(Mumberg et al., 1995)
pEI091	p415GPD-mNeonGreen-Myc-GSGSGS-SceI$^{cut\ site}$-FCY1-SceI$^{cut\ site}$-T8term	this study
pEI080	p413TEF-NLS-SCEI	this study
pEI087	p415GPD-mNeonGreen-p415GPD-mNeonGreen-SceIsite-FCY1-SceIsite-T8term	this study
pYM-N19	pFA6a-natNT2-TEFpr	(Janke et al., 2004)
pYM-N7	pFA6a-natNT2-ADHpr	(Janke et al., 2004)

Table S 3. Localization of yeast native TA proteins.

ORF	Protein	Tagging	Localization	
			mNG-ORF	GPDpr-mNG-ORF
YBL091C-A	Scs22	N-SWAT	below threshold	cytosol
YLR026C	Sed5	N-SWAT	Golgi	Golgi
YBR162W-A	Ysy6	N-SWAT	ER	ER
YDR086C	Sss1	N-SWAT	ER	ER
YDR200C	Far9/Vps64	N-SWAT	below threshold	ER
YDR498C	Sec20	N-SWAT	ER	ER
YER087C-B	Sbh1	N-SWAT	ER	ER
YER100W	Ubc6	N-SWAT	ER	ER
YER120W	Scs2	N-SWAT	ER	ER
YGL098W	Use1	N-SWAT	ER	ER
YLR238W	Far10	N-SWAT	below threshold	ER
YLR268W	Sec22	N-SWAT	ER	ER
YMR161W	Hlj1	N-SWAT	ER	ER
YNL111C	Cyb5	N-SWAT	ER	ER
YNL188W	Kar1	N-SWAT	ER	ER
YOR075W	Ufe1	N-SWAT	below threshold	ER
YOR324C	Frt1	N-SWAT	below threshold	ER
YPR183W	Dpm1	N-SWAT	below threshold	ER
YDR468C	Tlg1	N-SWAT	Golgi	Golgi
YHL031C	Gos1	N-SWAT	Golgi	Golgi
YIL004C	Bet1	N-SWAT	below threshold	Golgi
YKL006C-A	Sft1	N-SWAT	below threshold	Golgi
YLR078C	Bos1	N-SWAT	Golgi	Golgi
YMR197C	Vti1	N-SWAT	below threshold	Golgi
YOL018C	Tlg2	N-SWAT	below threshold	Golgi
YFL046W	Fmp32	N-SWAT	below threshold	mitochondria

YIL065C	Fis1	N-SWAT	mitochondria	mitochondria
YNL070W	Tom7	N-SWAT	mitochondria	mitochondria
YNL131W	Tom22	N-SWAT	mitochondria	mitochondria
YOR045W	Tom6	N-SWAT	mitochondria	mitochondria
YPR133W-A	Tom5	N-SWAT	mitochondria	mitochondria
YDL012C	Ydl012c	N-SWAT	PM/bud	Golgi
YPL192C	Prm3	N-SWAT	below threshold	NE
YOL044W	Pex15	N-SWAT	peroxisome	peroxisome
YAL030W	Snc1	N-SWAT	below threshold	puncta
YDR034W-B	Ydr034w-b	N-SWAT	PM	PM
YDR210W	Ydr210w	N-SWAT	PM	PM
YMR183C	Sso2	N-SWAT	PM	PM
YOR327C	Snc2	N-SWAT	Golgi	PM/bud
YPL232W	Sso1	N-SWAT	PM	PM
YDL241W	Ydl241w	N-SWAT	below threshold	puncta
YDR281C	Phm6	N-SWAT	below threshold	vacuole
YLR093C	Nyv1	N-SWAT	vacuole	vacuole
YOR036W	Pep12	N-SWAT	Golgi	vacuole
YOR106W	Vam3	N-SWAT	vacuole, low signal	vacuole
YPL200W	Csm4	N-SWAT	below threshold	cytosol
YBL100C	Ybl100c	PCR targeting	NA	ER
YAL028W	Frt2	N-SWAT	below threshold	ER
YDR522C	Sps2	N-SWAT	below threshold	cytosol
YEL073C	Yel073c	N-SWAT	below threshold	cytosol, puncta
YML036W	Cgi121	N-SWAT	below threshold	cytosol
YER019C-A	Sbh2	PCR targeting	NA	ER
YAL014C	Syn8	PCR targeting	below threshold	Golgi
YBR016W	Ybr016w	PCR targeting	PM	PM
YBR056W-A	Mnc1	PCR targeting	PM/cytosol	PM/cytosol
YPL206C	Pgc1	PCR targeting	NA	ER/puncta
YJL119C	Yjl119c	PCR targeting	NA	puncta
YEL010W	Yel010w	NA	NA	NA
YCL007C	Ycl007c	NA	NA	NA

Table S 4. Oligonucleotide pools encoding TMD variants.

Pool	Variant	Sequence
1	YAL014c/Syn8:TAvariants_1_1_	ggttctggatcaggttcaAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTATTAtaactccaggagTATATAAACTCAT
1	YAL014c/Syn8:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTATTAtaactccaggagTATATAAACTCAT
1	YAL014c/Syn8:TAvariants_5_1_	ggttctggatcaggttcaAGGTTCAAGGATAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTATTAtaactccaggagTATATAAACTCAT
1	YAL028w/Frt2:TAvariants_1_1_	ggttctggatcaggttcaGAGGATGACAATAAGGCGTTTATTAATGAATTAAAGATATCAGTCGAAGAAAGTGTAGCGCAATTACAAtaactccaggagTATATAAACTCAT
1	YAL028w/Frt2:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAGGATGACAATAAGGCGTTTATTAATGAATTAAAGATATCAGTCGAAGAAAGTGTAGCGCAATTACAAtaactccaggagTATATAAACTCAT
1	YAL028w/Frt2:TAvariants_1_5_	ggttctggatcaggttcaGAGGATGACAATAAGGCGTTTATTAATGAATTAAAGATATCAGTCGAAGAAAGTGTAGCGCAATTACAAGGATTGGAAAGGtaactccaggagTATATAAACTCAT
1	YAL028w/Frt2:TAvariants_5_1_	ggttctggatcaggttcaTCAGACTTTGACGAGGATGACAATAAGGCGTTTATTAATGAATTAAAGATATCAGTCGAAGAAAGTGTAGCGCAATTACAAtaactccaggagTATATAAACTCAT
1	YAL030w/Snc1:TAvariants_1_1_	ggttctggatcaggttcaATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTtaactccaggagTATATAAACTCAT
1	YAL030w/Snc1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTtaactccaggagTATATAAACTCAT
1	YAL030w/Snc1:TAvariants_1_5_	ggttctggatcaggttcaATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTCGAtaactccaggagTATATAAACTCAT
1	YAL030w/Snc1:TAvariants_5_1_	ggttctggatcaggttcaCTAAAAATGAAGATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTtaactccaggagTATATAAACTCAT
1	YBL091c-A/Scs22:TAvariants_1_1_	ggttctggatcaggttcaAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACTACtaactccaggagTATATAAACTCAT
1	YBL091c-A/Scs22:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACTACtaactccaggagTATATAAACTCAT
1	YBL091c-A/Scs22:TAvariants_5_1_	ggttctggatcaggttcaGGTAACGGGCAAAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACTACtaactccaggagTATATAAACTCAT
1	YBL100c:TAvariants_1_1_	ggttctggatcaggttcaATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTtaactccaggagTATATAAACTCAT
1	YBL100c:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTtaactccaggagTATATAAACTCAT
1	YBL100c:TAvariants_1_5_	ggttctggatcaggttcaATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTTACTGTTCTTGCtaactccaggagTATATAAACTCAT
1	YBL100c:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTTACTGTTCTTGCtaactccaggagTATATAAACTCAT

1	YBL100c:TAvariants_5_1_	ggttctggatcaggttcaGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTtaactccaggagTATATAAACTCAT
1	YBL100c:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTtaactccaggagTATATAAACTCAT
1	YBL100c:TAvariants_5_5_	ggttctggatcaggttcaGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTTACTGTTCTTGCtaactccaggagTATATAAACTCAT
1	YBR016w:TAvariants_1_1_	ggttctggatcaggttcaGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
1	YBR016w:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
1	YBR016w:TAvariants_5_1_	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
1	YBR162w-A/Ysy6:TAvariants_1_1_	ggttctggatcaggttcaACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCCTAtaactccaggagTATATAAACTCAT
1	YBR162w-A/Ysy6:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCCTAtaactccaggagTATATAAACTCAT
1	YBR162w-A/Ysy6:TAvariants_5_1_	ggttctggatcaggttcaGTGATATCCAAAACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCCTAtaactccaggagTATATAAACTCAT
1	YCL007c/Cwh36:TAvariants_1_1_	ggttctggatcaggttcaGACACATACAAACTACTCATTACTCTTTGTTCTTTATTATTCGTTGGACCTTTGTTTCTTAAAtaactccaggagTATATAAACTCAT
1	YCL007c/Cwh36:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGACACATACAAACTACTCATTACTCTTTGTTCTTTATTATTCGTTGGACCTTTGTTTCTTAAAtaactccaggagTATATAAACTCAT
1	YCL007c/Cwh36:TAvariants_1_5_	ggttctggatcaggttcaGACACATACAAACTACTCATTACTCTTTGTTCTTTATTATTCGTTGGACCTTTGTTTCTTAAAGTAtaactccaggagTATATAAACTCAT
1	YCL007c/Cwh36:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGACACATACAAACTACTCATTACTCTTTGTTCTTTATTATTCGTTGGACCTTTGTTTCTTAAAGTAtaactccaggagTATATAAACTCAT
1	YCL007c/Cwh36:TAvariants_5_1_	ggttctggatcaggttcaAGAAGAGATTATGACACATACAAACTACTCATTACTCTTTGTTCTTTATTATTCGTTGGACCTTTGTTTCTTAAAtaactccaggagTATATAAACTCAT
1	YCL007c/Cwh36:TAvariants_5_5_	ggttctggatcaggttcaAGAAGAGATTATGACACATACAAACTACTCATTACTCTTTGTTCTTTATTATTCGTTGGACCTTTGTTTCTTAAAGTAtaactccaggagTATATAAACTCAT
1	YDL012c:TAvariants_1_1_	ggttctggatcaggttcaGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccaggagTATATAAACTCAT
1	YDL012c:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccaggagTATATAAACTCAT
1	YDL012c:TAvariants_5_1_	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccaggagTATATAAACTCAT
1	YDL241w:TAvariants_1_1_	ggttctggatcaggttcaTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCtaactccaggagTATATAAACTCAT
1	YDL241w:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCtaactccaggagTATATAAACTCAT
1	YDL241w:TAvariants_5_1_	ggttctggatcaggttcaAGGCCTTCTCATTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCtaactccaggagTATATAAACTCAT
1	YDL241w:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGGCCTTCTCATTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCtaactccaggagTATATAAACTCAT
1	YDR034W-B:TAvariants_1_1_	ggttctggatcaggttcaGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACtaactccaggagTATATAAACTCAT
1	YDR034W-B:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACtaactccaggagTATATAAACTCAT
1	YDR034W-B:TAvariants_1_5_	ggttctggatcaggttcaGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT
1	YDR034W-B:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT
1	YDR034W-B:TAvariants_5_1_	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACtaactccaggagTATATAAACTCAT
1	YDR034W-B:TAvariants_5_5_	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT
1	YDR086c/Sss1:TAvariants_1_1_	ggttctggatcaggttcaAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGtaactccaggagTATATAAACTCAT
1	YDR086c/Sss1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGtaactccaggagTATATAAACTCAT
1	YDR086c/Sss1:TAvariants_1_5_	ggttctggatcaggttcaAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGTTGATTCATATTtaactccaggagTATATAAACTCAT
1	YDR086c/Sss1:TAvariants_5_1_	ggttctggatcaggttcaAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGtaactccaggagTATATAAACTCAT
1	YDR200c/Far9:TAvariants_1_1_	ggttctggatcaggttcaAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGtaactccaggagTATATAAACTCAT
1	YDR200c/Far9:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGtaactccaggagTATATAAACTCAT
1	YDR200c/Far9:TAvariants_1_5_	ggttctggatcaggttcaAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGCAATTGCCACAAtaactccaggagTATATAAACTCAT
1	YDR200c/Far9:TAvariants_5_1_	ggttctggatcaggttcaACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGtaactccaggagTATATAAACTCAT
1	YDR210w:TAvariants_1_1_	ggttctggatcaggttcaAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAAtaactccaggagTATATAAACTCAT
1	YDR210w:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAAtaactccaggagTATATAAACTCAT
1	YDR210w:TAvariants_1_5_	ggttctggatcaggttcaAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
1	YDR210w:TAvariants_5_1_	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAAtaactccaggagTATATAAACTCAT
1	YDR281c/Phm6:TAvariants_1_1_	ggttctggatcaggttcaGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGtaactccaggagTATATAAACTCAT
1	YDR468c/Tlg1:TAvariants_1_1_	ggttctggatcaggttcaGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGCTtaactccaggagTATATAAACTCAT
1	YDR468c/Tlg1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGCTtaactccaggagTATATAAACTCAT
1	YDR468c/Tlg1:TAvariants_5_1_	ggttctggatcaggttcaGAAAAATACGACGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGCTtaactccaggagTATATAAACTCAT
1	YDR498c/Sec20:TAvariants_1_1_	ggttctggatcaggttcaGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTtaactccaggagTATATAAACTCAT
1	YDR498c/Sec20:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTtaactccaggagTATATAAACTCAT

1	YDR498c/Sec20:TAvariants_1_5_	ggttctggatcaggttcaGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTCGTATTTTCAAGtaactccaggagTATATAAACTCAT
1	YDR498c/Sec20:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTCGTATTTTCAAGtaactccaggagTATATAAACTCAT
1	YDR498c/Sec20:TAvariants_5_1_	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTtaactccaggagTATATAAACTCAT
1	YDR498c/Sec20:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTtaactccaggagTATATAAACTCAT
1	YDR498c/Sec20:TAvariants_5_5_	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTCGTATTTTCAAGtaactccaggagTATATAAACTCAT
1	YDR522c/Sps2:TAvariants_1_1_	ggttctggatcaggttcaAACATTTTTATTGATGCTTTCAAAATGTCAGTTTATGCAGTTTTCACGGTATTGTTCTCGATAATTTTTtaactccaggagTATATAAACTCAT
1	YDR522c/Sps2:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAACATTTTTATTGATGCTTTCAAAATGTCAGTTTATGCAGTTTTCACGGTATTGTTCTCGATAATTTTTtaactccaggagTATATAAACTCAT
1	YDR522c/Sps2:TAvariants_5_1_	ggttctggatcaggttcaAGTGCTCCGAAGAACATTTTTATTGATGCTTTCAAAATGTCAGTTTATGCAGTTTTCACGGTATTGTTCTCGATAATTTTTtaactccaggagTATATAAACTCAT
1	YEL010w:TAvariants_1_1_	ggttctggatcaggttcaAGGCTTTCGAGCTTTTCTATTATTTTAAGTGCTGTGTTTCCGGACGTGCTCTTCACTTTCTTATTTTCTtaactccaggagTATATAAACTCAT
1	YEL010w:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGGCTTTCGAGCTTTTCTATTATTTTAAGTGCTGTGTTTCCGGACGTGCTCTTCACTTTCTTATTTTCTtaactccaggagTATATAAACTCAT
1	YEL010w:TAvariants_5_1_	ggttctggatcaggttcaAGGTTAGAGACGAGGCTTTCGAGCTTTTCTATTATTTTAAGTGCTGTGTTTCCGGACGTGCTCTTCACTTTCTTATTTTCTtaactccaggagTATATAAACTCAT
1	YEL073c:TAvariants_1_1_	ggttctggatcaggttcaGAAACTTCAGGTTATTATGCCTTTTCAACAGTAGTACCTGTTTTAATGGGGAATATGAAAGTAGCAtaactccaggagTATATAAACTCAT
1	YEL073c:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAAACTTCAGGTTATTATGCCTTTTCAACAGTAGTACCTGTTTTAATGGGGAATATGAAAGTAGCAtaactccaggagTATATAAACTCAT
1	YEL073c:TAvariants_5_1_	ggttctggatcaggttcaACTAACCCATACGAAACTTCAGGTTATTATGCCTTTTCAACAGTAGTACCTGTTTTAATGGGGAATATGAAAGTAGCAtaactccaggagTATATAAACTCAT
1	YER019c-A/Sbh2:TAvariants_1_1_	ggttctggatcaggttcaTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGtaactccaggagTATATAAACTCAT
1	YER019c-A/Sbh2:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGtaactccaggagTATATAAACTCAT
1	YER019c-A/Sbh2:TAvariants_1_5_	ggttctggatcaggttcaTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGAAATTTACACACtaactccaggagTATATAAACTCAT
1	YER019c-A/Sbh2:TAvariants_5_1_	ggttctggatcaggttcaTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGtaactccaggagTATATAAACTCAT
1	YER087c-B/Sbh1:TAvariants_1_1_	ggttctggatcaggttcaCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTtaactccaggagTATATAAACTCAT
1	YER087c-B/Sbh1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTtaactccaggagTATATAAACTCAT
1	YER087c-B/Sbh1:TAvariants_1_5_	ggttctggatcaggttcaCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTAAAGTTGCCGGTtaactccaggagTATATAAACTCAT
1	YER087c-B/Sbh1:TAvariants_5_1_	ggttctggatcaggttcaCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTtaactccaggagTATATAAACTCAT
1	YER100w/Ubc6:TAvariants_1_1_	ggttctggatcaggttcaTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGAAAtaactccaggagTATATAAACTCAT
1	YER100w/Ubc6:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGAAAtaactccaggagTATATAAACTCAT
1	YER100w/Ubc6:TAvariants_5_1_	ggttctggatcaggttcaCCTAATGATAGTTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGAAAtaactccaggagTATATAAACTCAT
1	YER120w/Scs2:TAvariants_1_1_	ggttctggatcaggttcaAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACAGAtaactccaggagTATATAAACTCAT
1	YER120w/Scs2:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACAGAtaactccaggagTATATAAACTCAT
1	YER120w/Scs2:TAvariants_5_1_	ggttctggatcaggttcaAATGAATCATCCAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACAGAtaactccaggagTATATAAACTCAT
1	YFL046w/Fmp32:TAvariants_1_1_	ggttctggatcaggttcaCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGtaactccaggagTATATAAACTCAT
1	YFL046w/Fmp32:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGtaactccaggagTATATAAACTCAT
1	YFL046w/Fmp32:TAvariants_1_5_	ggttctggatcaggttcaCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGTTATTAACAtaactccaggagTATATAAACTCAT
1	YFL046w/Fmp32:TAvariants_5_1_	ggttctggatcaggttcaTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGtaactccaggagTATATAAACTCAT
1	YGL098w/Use1:TAvariants_1_1_	ggttctggatcaggttcaAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGtaactccaggagTATATAAACTCAT
1	YHL031c/Gos1:TAvariants_1_1_	ggttctggatcaggttcaAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACATGGtaactccaggagTATATAAACTCAT
1	YHL031c/Gos1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACATGGtaactccaggagTATATAAACTCAT
1	YHL031c/Gos1:TAvariants_5_1_	ggttctggatcaggttcaAGAAGGAAGAAAAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACATGGtaactccaggagTATATAAACTCAT
1	YIL004c/Bet1:TAvariants_1_1_	ggttctggatcaggttcaGGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTACAtaactccaggagTATATAAACTCAT
1	YIL065c/Fis1:TAvariants_1_1_	ggttctggatcaggttcaAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAtaactccaggagTATATAAACTCAT
1	YIL065c/Fis1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAtaactccaggagTATATAAACTCAT
1	YIL065c/Fis1:TAvariants_1_5_	ggttctggatcaggttcaAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAAACAAGAGAAGGtaactccaggagTATATAAACTCAT
1	YIL065c/Fis1:TAvariants_5_1_	ggttctggatcaggttcaAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAtaactccaggagTATATAAACTCAT
1	YJL119c:TAvariants_1_1_	ggttctggatcaggttcaGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTtaactccaggagTATATAAACTCAT
1	YJL119c:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTtaactccaggagTATATAAACTCAT
1	YJL119c:TAvariants_1_5_	ggttctggatcaggttcaGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTTTTTTCGCAtaactccaggagTATATAAACTCAT
1	YJL119c:TAvariants_5_1_	ggttctggatcaggttcaACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTtaactccaggagTATATAAACTCAT
1	YKL006c-A/Sft1:TAvariants_1_1_	ggttctggatcaggttcaAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGtaactccaggagTATATAAACTCAT

1	YKL006c-A/Sft1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGtaactccaggagTATATAAACTCAT
1	YKL006c-A/Sft1:TAvariants_1_5_	ggttctggatcaggttcaAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGTTATTTtaactccaggagTATATAAACTCAT
1	YKL006c-A/Sft1:TAvariants_5_1_	ggttctggatcaggttcaTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGtaactccaggagTATATAAACTCAT
1	YLR026c/Sed5:TAvariants_1_1_	ggttctggatcaggttcaAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCAATtaactccaggagTATATAAACTCAT
1	YLR026c/Sed5:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCAATtaactccaggagTATATAAACTCAT
1	YLR026c/Sed5:TAvariants_5_1_	ggttctggatcaggttcaATAAAGAGTAATAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCAATtaactccaggagTATATAAACTCAT
1	YLR078c/Bos1:TAvariants_1_1_	ggttctggatcaggttcaAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAAtaactccaggagTATATAAACTCAT
1	YLR078c/Bos1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAAtaactccaggagTATATAAACTCAT
1	YLR078c/Bos1:TAvariants_1_5_	ggttctggatcaggttcaAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAATGGTTAAGAtaactccaggagTATATAAACTCAT
1	YLR078c/Bos1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAATGGTTAAGAtaactccaggagTATATAAACTCAT
1	YLR078c/Bos1:TAvariants_5_1_	ggttctggatcaggttcaGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAATaactccaggagTATATAAACTCAT
1	YLR078c/Bos1:TAvariants_5_5_	ggttctggatcaggttcaGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAATGGTTAAGAtaactccaggagTATATAAACTCAT
1	YLR093c/Nyv1:TAvariants_1_1_	ggttctggatcaggttcaAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGTGGtaactccaggagTATATAAACTCAT
1	YLR093c/Nyv1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGTGGtaactccaggagTATATAAACTCAT
1	YLR093c/Nyv1:TAvariants_5_1_	ggttctggatcaggttcaCAGAAGGTCAAAAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGTGGtaactccaggagTATATAAACTCAT
1	YLR238w/Far10:TAvariants_1_1_	ggttctggatcaggttcaCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCtaactccaggagTATATAAACTCAT
1	YLR238w/Far10:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCtaactccaggagTATATAAACTCAT
1	YLR238w/Far10:TAvariants_1_5_	ggttctggatcaggttcaCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCCCCAACtaactccaggagTATATAAACTCAT
1	YLR238w/Far10:TAvariants_5_1_	ggttctggatcaggttcaACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCtaactccaggagTATATAAACTCAT
1	YML036w/Cgi121:TAvariants_1_1_	ggttctggatcaggttcaGATGATTTCAAGCCCCAAGACGTAAATGGTCTCTCAAGAGCTTTGGTAGACGCTATTCAATTGAGGGGTGTGtaactccaggagTATATAAACTCAT
1	YMR161w/Hlj1:TAvariants_1_1_	ggttctggatcaggttcaAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTAGTtaactccaggagTATATAAACTCAT
1	YMR161w/Hlj1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTAGTtaactccaggagTATATAAACTCAT
1	YMR161w/Hlj1:TAvariants_5_1_	ggttctggatcaggttcaTCGCAATTAAAAAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTAGTtaactccaggagTATATAAACTCAT
1	YMR183c/Sso2:TAvariants_1_1_	ggttctggatcaggttcaAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTGGAAtaactccaggagTATATAAACTCAT
1	YMR197c/Vti1:TAvariants_1_1_	ggttctggatcaggttcaAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAtaactccaggagTATATAAACTCAT
1	YMR197c/Vti1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAtaactccaggagTATATAAACTCAT
1	YMR197c/Vti1:TAvariants_1_5_	ggttctggatcaggttcaAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAAAGTTTAAAtaactccaggagTATATAAACTCAT
1	YMR197c/Vti1:TAvariants_5_1_	ggttctggatcaggttcaCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAtaactccaggagTATATAAACTCAT
1	YNL070w/Tom7:TAvariants_1_1_	ggttctggatcaggttcaAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGCAtaactccaggagTATATAAACTCAT
1	YNL111c/Cyb5:TAvariants_1_1_	ggttctggatcaggttcaAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACtaactccaggagTATATAAACTCAT
1	YNL111c/Cyb5:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACtaactccaggagTATATAAACTCAT
1	YNL111c/Cyb5:TAvariants_1_5_	ggttctggatcaggttcaAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACGAAtaactccaggagTATATAAACTCAT
1	YNL111c/Cyb5:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACGAAtaactccaggagTATATAAACTCAT
1	YNL111c/Cyb5:TAvariants_5_1_	ggttctggatcaggttcaCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACtaactccaggagTATATAAACTCAT
1	YNL111c/Cyb5:TAvariants_5_5_	ggttctggatcaggttcaCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACGAAtaactccaggagTATATAAACTCAT
1	YNL131w/Tom22:TAvariants_1_1_	ggttctggatcaggttcaAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGAAtaactccaggagTATATAAACTCAT
1	YNL188w/Kar1:TAvariants_1_1_	ggttctggatcaggttcaGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGGtaactccaggagTATATAAACTCAT
1	YNL188w/Kar1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGGtaactccaggagTATATAAACTCAT
1	YNL188w/Kar1:TAvariants_1_5_	ggttctggatcaggttcaGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGGTTTtaactccaggagTATATAAACTCAT
1	YNL188w/Kar1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGGTTTtaactccaggagTATATAAACTCAT
1	YNL188w/Kar1:TAvariants_5_1_	ggttctggatcaggttcaAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGGtaactccaggagTATATAAACTCAT
1	YNL188w/Kar1:TAvariants_5_5_	ggttctggatcaggttcaAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGGTTTtaactccaggagTATATAAACTCAT
1	YOL018c/Tlg2:TAvariants_1_1_	ggttctggatcaggttcaAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAAtaactccaggagTATATAAACTCAT
1	YOL018c/Tlg2:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAAtaactccaggagTATATAAACTCAT
1	YOL018c/Tlg2:TAvariants_1_5_	ggttctggatcaggttcaAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAACCACATGGCGGTtaactccaggagTATATAAACTCAT
1	YOL018c/Tlg2:TAvariants_5_1_	ggttctggatcaggttcaACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAAtaactccaggagTATATAAACTCAT

1	YOL044w/Pex15:TAvariants_1_1_	ggttctggatcaggttcaAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAtaactccaggagTATATAAACTCAT
1	YOL044w/Pex15:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAtaactccaggagTATATAAACTCAT
1	YOL044w/Pex15:TAvariants_1_5_	ggttctggatcaggttcaAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAAAATATAAGTCAtaactccaggagTATATAAACTCAT
1	YOL044w/Pex15:TAvariants_5_1_	ggttctggatcaggttcaCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAtaactccaggagTATATAAACTCAT
1	YOR036w/Pep12:TAvariants_1_1_	ggttctggatcaggttcaAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAAAtaactccaggagTATATAAACTCAT
1	YOR036w/Pep12:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAAAtaactccaggagTATATAAACTCAT
1	YOR036w/Pep12:TAvariants_1_5_	ggttctggatcaggttcaAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAAATTGtaactccaggagTATATAAACTCAT
1	YOR036w/Pep12:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAAATTGtaactccaggagTATATAAACTCAT
1	YOR036w/Pep12:TAvariants_5_1_	ggttctggatcaggttcaAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAAAtaactccaggagTATATAAACTCAT
1	YOR036w/Pep12:TAvariants_5_5_	ggttctggatcaggttcaAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAAATTGtaactccaggagTATATAAACTCAT
1	YOR045w/Tom6:TAvariants_1_1_	ggttctggatcaggttcaTCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAAtaactccaggagTATATAAACTCAT
1	YOR045w/Tom6:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAAtaactccaggagTATATAAACTCAT
1	YOR045w/Tom6:TAvariants_1_5_	ggttctggatcaggttcaTCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAATCTCCACTCATGtaactccaggagTATATAAACTCAT
1	YOR045w/Tom6:TAvariants_5_1_	ggttctggatcaggttcaGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAAtaactccaggagTATATAAACTCAT
1	YOR075w/Ufe1:TAvariants_1_1_	ggttctggatcaggttcaAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGGTtaactccaggagTATATAAACTCAT
1	YOR075w/Ufe1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGGTtaactccaggagTATATAAACTCAT
1	YOR075w/Ufe1:TAvariants_5_1_	ggttctggatcaggttcaGGAAGAACTGCTAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGGTtaactccaggagTATATAAACTCAT
1	YOR106w/Vam3:TAvariants_1_1_	ggttctggatcaggttcaAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAAGTtaactccaggagTATATAAACTCAT
1	YOR106w/Vam3:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAAGTtaactccaggagTATATAAACTCAT
1	YOR106w/Vam3:TAvariants_5_1_	ggttctggatcaggttcaAACAAATGCGGTAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAAGTtaactccaggagTATATAAACTCAT
1	YOR324c/Frt1:TAvariants_1_1_	ggttctggatcaggttcaTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGtaactccaggagTATATAAACTCAT
1	YOR327c/Snc2:TAvariants_1_1_	ggttctggatcaggttcaAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCAGCtaactccaggagTATATAAACTCAT
1	YPL192c/Prm3:TAvariants_1_1_	ggttctggatcaggttcaTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAtaactccaggagTATATAAACTCAT
1	YPL200w/Csm4:TAvariants_1_1_	ggttctggatcaggttcaGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAtaactccaggagTATATAAACTCAT
1	YPL200w/Csm4:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAtaactccaggagTATATAAACTCAT
1	YPL200w/Csm4:TAvariants_1_5_	ggttctggatcaggttcaGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAGTTtaactccaggagTATATAAACTCAT
1	YPL200w/Csm4:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAGTTtaactccaggagTATATAAACTCAT
1	YPL200w/Csm4:TAvariants_5_1_	ggttctggatcaggttcaTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAtaactccaggagTATATAAACTCAT
1	YPL200w/Csm4:TAvariants_5_5_	ggttctggatcaggttcaTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAGTTtaactccaggagTATATAAACTCAT
1	YPL206c/Pgc1:TAvariants_1_1_	ggttctggatcaggttcaAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAtaactccaggagTATATAAACTCAT
1	YPL206c/Pgc1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAtaactccaggagTATATAAACTCAT
1	YPL206c/Pgc1:TAvariants_1_5_	ggttctggatcaggttcaAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAACCATTCATTTTtaactccaggagTATATAAACTCAT
1	YPL206c/Pgc1:TAvariants_5_1_	ggttctggatcaggttcaCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAtaactccaggagTATATAAACTCAT
1	YPL232w/Sso1:TAvariants_1_1_	ggttctggatcaggttcaAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAtaactccaggagTATATAAACTCAT
1	YPL232w/Sso1:TAvariants_1_5_	ggttctggatcaggttcaAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAACGCGTtaactccaggagTATATAAACTCAT
1	YPR133w-A/Tom5:TAvariants_1_1_	ggttctggatcaggttcaCAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAtaactccaggagTATATAAACTCAT
1	YPR133w-A/Tom5:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAtaactccaggagTATATAAACTCAT
1	YPR133w-A/Tom5:TAvariants_1_5_	ggttctggatcaggttcaCAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAAAGCAATGGAAAtaactccaggagTATATAAACTCAT
1	YPR133w-A/Tom5:TAvariants_5_1_	ggttctggatcaggttcaAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAtaactccaggagTATATAAACTCAT
1	YPR183w/Dpm1:TAvariants_1_1_	ggttctggatcaggttcaAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATtaactccaggagTATATAAACTCAT
1	YPR183w/Dpm1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATtaactccaggagTATATAAACTCAT
1	YPR183w/Dpm1:TAvariants_1_5_	ggttctggatcaggttcaAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATTTGGTCTTTtaactccaggagTATATAAACTCAT
1	YPR183w/Dpm1:TAvariants_5_1_	ggttctggatcaggttcaTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATtaactccaggagTATATAAACTCAT
2	YAL014c/Syn8:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGGTTCAAGGATAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTATTAtaactccaggagTATATAAACTCAT
2	YAL028w/Frt2:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAGGATGACAATAAGGCGTTTATTAATGAATTAAAGATATCAGTCGAAGAAAGTGTAGCGCAATTACAAGGATTGGAAAGGtaactccaggagTATATAAACTCAT

2	YAL028w/Frt2:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCAGACTTTGACGAGGATGACAATAAGGCGTTTATTAATGAATTAAAGATATCAGTCGAAGAAAGTGTAGCGCAATTACAAtaactccaggagTATATAAACTCAT
2	YAL028w/Frt2:TAvariants_5_5_	ggttctggatcaggttcaTCAGACTTTGACGAGGATGACAATAAGGCGTTTATTAATGAATTAAAGATATCAGTCGAAGAAAGTGTAGCGCAATTACAAGGATTGGAAAGGtaactccaggagTATATAAACTCAT
2	YAL028w/Frt2:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCAGACTTTGACGAGGATGACAATAAGGCGTTTATTAATGAATTAAAGATATCAGTCGAAGAAAGTGTAGCGCAATTACAAGGATTGGAAAGGtaactccaggagTATATAAACTCAT
2	YAL030w/Snc1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTCGAataactccaggagTATATAAACTCAT
2	YAL030w/Snc1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCTAAAAATGAAGATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTtaactccaggagTATATAAACTCAT
2	YAL030w/Snc1:TAvariants_5_5_	ggttctggatcaggttcaCTAAAAATGAAGATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTCGAtaactccaggagTATATAAACTCAT
2	YAL030w/Snc1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCTAAAAATGAAGATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTCGAataactccaggagTATATAAACTCAT
2	YBL091c-A/Scs22:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGGTAACGGGCAAAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACTACTaactccaggagTATATAAACTCAT
2	YBL100c:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTTACTGTTTCTTGCtaactccaggagTATATAAACTCAT
2	YBR016w:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
2	YBR162w-A/Ysy6:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGTGATATCCAAAACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCCTAtaactccaggagTATATAAACTCAT
2	YCL007c/Cwh36:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGAAGAGATTATGACACATACAAACTACTCATTACTCTTTGTTCTTTATTATTCGTTGGACCTTTGTTTCTTAAAtaactccaggagTATATAAACTCAT
2	YCL007c/Cwh36:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGAAGAGATTATGACACATACAAACTACTCATTACTCTTTGTTCTTTATTATTCGTTGGACCTTTGTTTCTTAAAGTAtaactccaggagTATATAAACTCAT
2	YDL012c:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccaggagTATATAAACTCAT
2	YDR034W-B:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACtaactccaggagTATATAAACTCAT
2	YDR034W-B:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT
2	YDR086c/Sss1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGTTGATTCATATTtaactccaggagTATATAAACTCAT
2	YDR086c/Sss1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGtaactccaggagTATATAAACTCAT
2	YDR086c/Sss1:TAvariants_5_5_	ggttctggatcaggttcaAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGTTGATTCATATTtaactccaggagTATATAAACTCAT
2	YDR086c/Sss1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGTTGATTCATATTtaactccaggagTATATAAACTCAT
2	YDR200c/Far9:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGCAATTGCCACAAtaactccaggagTATATAAACTCAT
2	YDR200c/Far9:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGtaactccaggagTATATAAACTCAT
2	YDR200c/Far9:TAvariants_5_5_	ggttctggatcaggttcaACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGCAATTGCCACAAtaactccaggagTATATAAACTCAT
2	YDR200c/Far9:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGCAATTGCCACAAtaactccaggagTATATAAACTCAT
2	YDR210w:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
2	YDR210w:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAAtaactccaggagTATATAAACTCAT
2	YDR210w:TAvariants_5_5_	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
2	YDR210w:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
2	YDR281c/Phm6:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGtaactccaggagTATATAAACTCAT
2	YDR281c/Phm6:TAvariants_1_5_	ggttctggatcaggttcaGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGACGATGACGAAGtaactccaggagTATATAAACTCAT
2	YDR281c/Phm6:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGACGATGACGAAGtaactccaggagTATATAAACTCAT
2	YDR281c/Phm6:TAvariants_5_1_	ggttctggatcaggttcaGGGTTTCTCGAGGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGtaactccaggagTATATAAACTCAT
2	YDR281c/Phm6:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGGGTTTCTCGAGGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGtaactccaggagTATATAAACTCAT
2	YDR281c/Phm6:TAvariants_5_5_	ggttctggatcaggttcaGGGTTTCTCGAGGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGACGATGACGAAGtaactccaggagTATATAAACTCAT
2	YDR281c/Phm6:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGGGTTTCTCGAGGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGACGATGACGAAGtaactccaggagTATATAAACTCAT
2	YDR468c/Tlg1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAAAAATACGACGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGCTtaactccaggagTATATAAACTCAT
2	YDR498c/Sec20:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTCGTATTTTCAAGtaactccaggagTATATAAACTCAT
2	YDR522c/Sps2:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGTGCTCCGAAGAACATTTTTATTGATGCTTTCAAAATGTCAGTTTATGCAGTTTTCACGGTATTGTTCTCGATAATTTTTtaactccaggagTATATAAACTCAT
2	YEL010w:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGGTTAGAGACGAGGCTTTCGAGCTTTTCTATTATTTTAAGTGCTGTGTTTCCGGACGTGCTCTTCACTTTCTTATTTTCTtaactccaggagTATATAAACTCAT
2	YEL073c:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACTAACCCATACGAAACTTCAGGTTATTATGCCTTTTCAACAGTAGTACCTGTTTTAATGGGGAATATGAAAGTAGCAtaactccaggagTATATAAACTCAT
2	YER019c-A/Sbh2:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGAAATTTACACACTaactccaggagTATATAAACTCAT
2	YER019c-A/Sbh2:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGtaactccaggagTATATAAACTCAT
2	YER019c-A/Sbh2:TAvariants_5_5_	ggttctggatcaggttcaTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGAAATTTACACACTaactccaggagTATATAAACTCAT
2	YER019c-A/Sbh2:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGAAATTTACACACTaactccaggagTATATAAACTCAT
2	YER087c-B/Sbh1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTAAAGTTGCCGGTtaactccaggagTATATAAACTCAT

2	YER087c-B/Sbh1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTtaactccaggagTATATAAACTCAT
2	YER087c-B/Sbh1:TAvariants_5_5_	ggttctggatcaggttcaCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTAAAGTTGCCGGTtaactccaggagTATATAAACTCAT
2	YER087c-B/Sbh1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTAAAGTTGCCGGTtaactccaggagTATATAAACTCAT
2	YER100w/Ubc6:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCCTAATGATAGTTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGAAAtaactccaggagTATATAAACTCAT
2	YER120w/Scs2:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAATGAATCATCCAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACAGAtaactccaggagTATATAAACTCAT
2	YFL046w/Fmp32:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGTTATTAACAtaactccaggagTATATAAACTCAT
2	YFL046w/Fmp32:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGtaactccaggagTATATAAACTCAT
2	YFL046w/Fmp32:TAvariants_5_5_	ggttctggatcaggttcaTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGTTATTAACAtaactccaggagTATATAAACTCAT
2	YFL046w/Fmp32:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGTTATTAACAtaactccaggagTATATAAACTCAT
2	YGL098w/Use1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGtaactccaggagTATATAAACTCAT
2	YGL098w/Use1:TAvariants_1_5_	ggttctggatcaggttcaAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGGCCCTAtaactccaggagTATATAAACTCAT
2	YGL098w/Use1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGGCCCTAtaactccaggagTATATAAACTCAT
2	YGL098w/Use1:TAvariants_5_1_	ggttctggatcaggttcaAAGAGTAAATTGAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGtaactccaggagTATATAAACTCAT
2	YGL098w/Use1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGAGTAAATTGAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGtaactccaggagTATATAAACTCAT
2	YGL098w/Use1:TAvariants_5_5_	ggttctggatcaggttcaAAGAGTAAATTGAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGGCCCTAtaactccaggagTATATAAACTCAT
2	YGL098w/Use1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGAGTAAATTGAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGGCCCTAtaactccaggagTATATAAACTCAT
2	YHL031c/Gos1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGAAGGAAGAAAAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACATGGtaactccaggagTATATAAACTCAT
2	YIL004c/Bet1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTACAtaactccaggagTATATAAACTCAT
2	YIL004c/Bet1:TAvariants_5_1_	ggttctggatcaggttcaGCCAGAAGATCTGGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTACAtaactccaggagTATATAAACTCAT
2	YIL004c/Bet1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGCCAGAAGATCTGGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTACAtaactccaggagTATATAAACTCAT
2	YIL065c/Fis1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAAACAAGAGAAGGtaactccaggagTATATAAACTCAT
2	YIL065c/Fis1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAtaactccaggagTATATAAACTCAT
2	YIL065c/Fis1:TAvariants_5_5_	ggttctggatcaggttcaAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAAACAAGAGAAGGtaactccaggagTATATAAACTCAT
2	YIL065c/Fis1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAAACAAGAGAAGGtaactccaggagTATATAAACTCAT
2	YJL119c:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTTTTTTCGCAtaactccaggagTATATAAACTCAT
2	YJL119c:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTtaactccaggagTATATAAACTCAT
2	YJL119c:TAvariants_5_5_	ggttctggatcaggttcaACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTTTTTTCGCAtaactccaggagTATATAAACTCAT
2	YJL119c:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTTTTTTCGCAtaactccaggagTATATAAACTCAT
2	YKL006c-A/Sft1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGTTATTTtaactccaggagTATATAAACTCAT
2	YKL006c-A/Sft1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGtaactccaggagTATATAAACTCAT
2	YKL006c-A/Sft1:TAvariants_5_5_	ggttctggatcaggttcaTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGTTATTTtaactccaggagTATATAAACTCAT
2	YKL006c-A/Sft1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGTTATTTtaactccaggagTATATAAACTCAT
2	YLR026c/Sed5:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtATAAAGAGTAATAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCAATtaactccaggagTATATAAACTCAT
2	YLR078c/Bos1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAAtaactccaggagTATATAAACTCAT
2	YLR078c/Bos1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAATGGTTAAGAtaactccaggagTATATAAACTCAT
2	YLR093c/Nyv1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAGAAGGTCAAAAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGTGGtaactccaggagTATATAAACTCAT
2	YLR238w/Far10:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCCCCAACtaactccaggagTATATAAACTCAT
2	YLR238w/Far10:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCtaactccaggagTATATAAACTCAT
2	YLR238w/Far10:TAvariants_5_5_	ggttctggatcaggttcaACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCCCCAACtaactccaggagTATATAAACTCAT
2	YLR238w/Far10:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCCCCAACtaactccaggagTATATAAACTCAT
2	YLR268w/Sec22:TAvariants_1_1_	ggttctggatcaggttcaAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCCTCtaactccaggagTATATAAACTCAT
2	YLR268w/Sec22:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCCTCtaactccaggagTATATAAACTCAT
2	YLR268w/Sec22:TAvariants_1_5_	ggttctggatcaggttcaAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCCTCAAAtaactccaggagTATATAAACTCAT
2	YLR268w/Sec22:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCCTCAAAtaactccaggagTATATAAACTCAT
2	YLR268w/Sec22:TAvariants_5_1_	ggttctggatcaggttcaGCGCAAAAGATCAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCCTCtaactccaggagTATATAAACTCAT
2	YLR268w/Sec22:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGCGCAAAAGATCAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCCTCtaactccaggagTATATAAACTCAT

2	YLR268w/Sec22:TAvariants_5_5_	ggttctggatcaggttcaGCGCAAAAGATCAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGT GGATCTTCCTCAAAtaactccaggagTATATAAACTCAT
2	YLR268w/Sec22:TAvariants_5_5_ggtg gaccaggt	ggttctggatcaggttcaggtggaccaggtGCGCAAAAGATCAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTT CTCTTCTGGTGGATCTTCCTCAAAtaactccaggagTATATAAACTCAT
2	YML036w/Cgi121:TAvariants_1_1_ggt ggaccaggt	ggttctggatcaggttcaggtggaccaggtGATGATTTCAAGCCCCAAGACGTAAATGGTCTCTCAAGAGCTTTGGTAGACGCTATTCAATTGAGGGGTGTG taactccaggagTATATAAACTCAT
2	YML036w/Cgi121:TAvariants_5_1_	ggttctggatcaggttcaTATAAATTAAGTGATGATTTCAAGCCCCAAGACGTAAATGGTCTCTCAAGAGCTTTGGTAGACGCTATTCAATTGAGGGGTG TGtaactccaggagTATATAAACTCAT
2	YML036w/Cgi121:TAvariants_5_1_ggt ggaccaggt	ggttctggatcaggttcaggtggaccaggtTATAAATTAAGTGATGATTTCAAGCCCCAAGACGTAAATGGTCTCTCAAGAGCTTTGGTAGACGCTATTCAAT TGAGGGGTGTGtaactccaggagTATATAAACTCAT
2	YMR161w/Hlj1:TAvariants_5_1_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtTCGCAATTAAAAAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTAGTt aactccaggagTATATAAACTCAT
2	YMR183c/Sso2:TAvariants_1_1_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTG GAAtaactccaggagTATATAAACTCAT
2	YMR183c/Sso2:TAvariants_1_5_	ggttctggatcaggttcaAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTGGAAACAAGA AAGtaactccaggagTATATAAACTCAT
2	YMR183c/Sso2:TAvariants_1_5_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTG GAAACAAGAAAGtaactccaggagTATATAAACTCAT
2	YMR183c/Sso2:TAvariants_5_1_	ggttctggatcaggttcaGCAAGAAAAAACAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTT GTGGAAtaactccaggagTATATAAACTCAT
2	YMR183c/Sso2:TAvariants_5_1_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtGCAAGAAAAAACAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGT TCCATCCGTTGTGGAAtaactccaggagTATATAAACTCAT
2	YMR183c/Sso2:TAvariants_5_5_	ggttctggatcaggttcaGCAAGAAAAAACAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTT GTGGAAACAAGAAAGtaactccaggagTATATAAACTCAT
2	YMR183c/Sso2:TAvariants_5_5_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtGCAAGAAAAAACAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGT TCCATCCGTTGTGGAAACAAGAAAGtaactccaggagTATATAAACTCAT
2	YMR197c/Vti1:TAvariants_1_5_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAAAGTTTAAAtaac tccaggagTATATAAACTCAT
2	YMR197c/Vti1:TAvariants_5_1_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCA taactccaggagTATATAAACTCAT
2	YMR197c/Vti1:TAvariants_5_5_	ggttctggatcaggttcaCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAAAGTTTAAAt aactccaggagTATATAAACTCAT
2	YMR197c/Vti1:TAvariants_5_5_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCA AAGTTTAAAtaactccaggagTATATAAACTCAT
2	YNL070w/Tom7:TAvariants_1_1_ggtg gaccaggt	ggttctggatcaggttcaggtggaccaggtAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGCAta actccaggagTATATAAACTCAT
2	YNL070w/Tom7:TAvariants_1_5_	ggttctggatcaggttcaAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGCACACACTTCTAAT taactccaggagTATATAAACTCAT
2	YNL070w/Tom7:TAvariants_1_5_ggtg gaccaggt	ggttctggatcaggttcaggtggaccaggtAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGCACA CACTTCTAATtaactccaggagTATATAAACTCAT
2	YNL070w/Tom7:TAvariants_5_1_	ggttctggatcaggttcaGAACGTATTTCCAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGC Ataactccaggag TATATAAACTCAT
2	YNL070w/Tom7:TAvariants_5_1_ggtg gaccaggt	ggttctggatcaggttcaggtggaccaggtGAACGTATTTCCAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTT GGGCTGGGCAtaactccaggagTATATAAACTCAT
2	YNL070w/Tom7:TAvariants_5_5_	ggttctggatcaggttcaGAACGTATTTCCAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGC ACACACTTCTAATtaactccaggagTATATAAACTCAT
2	YNL070w/Tom7:TAvariants_5_5_ggtg gaccaggt	ggttctggatcaggttcaggtggaccaggtGAACGTATTTCCAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTT GGGCTGGGCACACACTTCTAATtaactccaggagTATATAAACTCAT
2	YNL111c/Cyb5:TAvariants_5_1_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGA ACtaactccaggagTATATAAACTCAT
2	YNL111c/Cyb5:TAvariants_5_5_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGA ACGAAtaactccaggagTATATAAACTCAT
2	YNL131w/Tom22:TAvariants_1_1_ggt ggaccaggt	ggttctggatcaggttcaggtggaccaggtAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGAAtaa ctccaggagTATATAAACTCAT
2	YNL131w/Tom22:TAvariants_1_5_	ggttctggatcaggttcaAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGAACAACAGCTAATC taactccaggagTATATAAACTCAT
2	YNL131w/Tom22:TAvariants_1_5_ggt ggaccaggt	ggttctggatcaggttcaggtggaccaggtAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGAACA ACAGCTAATCtaactccaggagTATATAAACTCAT
2	YNL131w/Tom22:TAvariants_5_1_	ggttctggatcaggttcaACAAAATCCGGAAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGAA taactccaggagTATATAAACTCAT
2	YNL131w/Tom22:TAvariants_5_1_ggt ggaccaggt	ggttctggatcaggttcaggtggaccaggtACAAAATCCGGAAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTAT ACTTGCCGAAtaactccaggagTATATAAACTCAT
2	YNL131w/Tom22:TAvariants_5_5_	ggttctggatcaggttcaACAAAATCCGGAAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGAA CAACAGCTAATCtaactccaggagTATATAAACTCAT
2	YNL131w/Tom22:TAvariants_5_5_ggt ggaccaggt	ggttctggatcaggttcaggtggaccaggtACAAAATCCGGAAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTAT ACTTGCCGAACAACAGCTAATCtaactccaggagTATATAAACTCAT
2	YNL188w/Kar1:TAvariants_5_1_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTA TAGGtaactccaggagTATATAAACTCAT
2	YNL188w/Kar1:TAvariants_5_5_ggtgg accaggt	ggttctggatcaggttcaggtggaccaggtAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTA TAGGTTTtaactccaggagTATATAAACTCAT
2	YOL018c/Tlg2:TAvariants_1_5_ggtgga ccaggt	ggttctggatcaggttcaggtggaccaggtAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAACCACATGGCGGTtaa ctccaggagTATATAAACTCAT
2	YOL018c/Tlg2:TAvariants_5_1_ggtgga ccaggt	ggttctggatcaggttcaggtggaccaggtACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAAtaa ctccaggagTATATAAACTCAT
2	YOL018c/Tlg2:TAvariants_5_5_	ggttctggatcaggttcaACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAACCACATGGCGGT taactccaggagTATATAAACTCAT
2	YOL018c/Tlg2:TAvariants_5_5_ggtgga ccaggt	ggttctggatcaggttcaggtggaccaggtACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAACC ACATGGCGGTtaactccaggagTATATAAACTCAT
2	YOL044w/Pex15:TAvariants_1_5_ggtg gaccaggt	ggttctggatcaggttcaggtggaccaggtAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAAAATATAAGTCAt aactccaggagTATATAAACTCAT
2	YOL044w/Pex15:TAvariants_5_1_ggtg gaccaggt	ggttctggatcaggttcaggtggaccaggtCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAta actccaggagTATATAAACTCAT
2	YOL044w/Pex15:TAvariants_5_5_	ggttctggatcaggttcaCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAAAATATAAGTC Ataactccaggag TATATAAACTCAT
2	YOL044w/Pex15:TAvariants_5_5_ggtg gaccaggt	ggttctggatcaggttcaggtggaccaggtCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAA AATATAAGTCAtaactccaggagTATATAAACTCAT
2	YOR036w/Pep12:TAvariants_5_1_ggt ggaccaggt	ggttctggatcaggttcaggtggaccaggtAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTAT GAAAtaactccaggagTATATAAACTCAT
2	YOR036w/Pep12:TAvariants_5_5_ggt ggaccaggt	ggttctggatcaggttcaggtggaccaggtAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTAT GAAATTGtaactccaggagTATATAAACTCAT
2	YOR045w/Tom6:TAvariants_1_5_ggtg gaccaggt	ggttctggatcaggttcaggtggaccaggtTCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAATCTCCACTCAT GtaactccaggagTATATAAACTCAT

2	YOR045w/Tom6:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAAtaactccaggagTATATAAACTCAT
2	YOR045w/Tom6:TAvariants_5_5_	ggttctggatcaggttcaGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAATCTCCACTCATGtaactccaggagTATATAAACTCAT
2	YOR045w/Tom6:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAATCTCCACTCATGtaactccaggagTATATAAACTCAT
2	YOR075w/Ufe1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGGAAGAACTGCTAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGGTtaactccaggagTATATAAACTCAT
2	YOR106w/Vam3:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAACAAATGCGGTAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAAGTtaactccaggagTATATAAACTCAT
2	YOR324c/Frt1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGtaactccaggagTATATAAACTCAT
2	YOR324c/Frt1:TAvariants_1_5_	ggttctggatcaggttcaTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGAATTTGTTAACTtaactccaggagTATATAAACTCAT
2	YOR324c/Frt1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGAATTTGTTAACTtaactccaggagTATATAAACTCAT
2	YOR324c/Frt1:TAvariants_5_1_	ggttctggatcaggttcaACGAAATCCATCTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGtaactccaggagTATATAAACTCAT
2	YOR324c/Frt1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACGAAATCCATCTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGtaactccaggagTATATAAACTCAT
2	YOR324c/Frt1:TAvariants_5_5_	ggttctggatcaggttcaACGAAATCCATCTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGAATTTGTTAACTtaactccaggagTATATAAACTCAT
2	YOR324c/Frt1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtACGAAATCCATCTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGAATTTGTTAACTtaactccaggagTATATAAACTCAT
2	YOR327c/Snc2:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCAGCtaactccaggagTATATAAACTCAT
2	YOR327c/Snc2:TAvariants_5_1_	ggttctggatcaggttcaGATCTAAAAATGAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCAGCtaactccaggagTATATAAACTCAT
2	YOR327c/Snc2:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGATCTAAAAATGAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCAGCtaactccaggagTATATAAACTCAT
2	YPL192c/Prm3:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAtaactccaggagTATATAAACTCAT
2	YPL192c/Prm3:TAvariants_1_5_	ggttctggatcaggttcaTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAGCTCTTCAAAATtaactccaggagTATATAAACTCAT
2	YPL192c/Prm3:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAGCTCTTCAAAATtaactccaggagTATATAAACTCAT
2	YPL192c/Prm3:TAvariants_5_1_	ggttctggatcaggttcaGAGAATAAGGGTTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAtaactccaggagTATATAAACTCAT
2	YPL192c/Prm3:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAGAATAAGGGTTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAtaactccaggagTATATAAACTCAT
2	YPL192c/Prm3:TAvariants_5_5_	ggttctggatcaggttcaGAGAATAAGGGTTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAGCTCTTCAAAATtaactccaggagTATATAAACTCAT
2	YPL192c/Prm3:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtGAGAATAAGGGTTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAGCTCTTCAAAATtaactccaggagTATATAAACTCAT
2	YPL200w/Csm4:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAtaactccaggagTATATAAACTCAT
2	YPL200w/Csm4:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAGTTtaactccaggagTATATAAACTCAT
2	YPL206c/Pgc1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAACCATTCATTTTtaactccaggagTATATAAACTCAT
2	YPL206c/Pgc1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAtaactccaggagTATATAAACTCAT
2	YPL206c/Pgc1:TAvariants_5_5_	ggttctggatcaggttcaCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAACCATTCATTTTtaactccaggagTATATAAACTCAT
2	YPL206c/Pgc1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAACCATTCATTTTtaactccaggagTATATAAACTCAT
2	YPL232w/Sso1:TAvariants_1_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAtaactccaggagTATATAAACTCAT
2	YPL232w/Sso1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAACGCGTtaactccaggagTATATAAACTCAT
2	YPL232w/Sso1:TAvariants_5_1_	ggttctggatcaggttcaAAGAACAAGATTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAtaactccaggagTATATAAACTCAT
2	YPL232w/Sso1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGAACAAGATTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAtaactccaggagTATATAAACTCAT
2	YPL232w/Sso1:TAvariants_5_5_	ggttctggatcaggttcaAAGAACAAGATTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAACGCGTtaactccaggagTATATAAACTCAT
2	YPL232w/Sso1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAGAACAAGATTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAACGCGTtaactccaggagTATATAAACTCAT
2	YPR133w-A/Tom5:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtCAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAAAGCAATGGAAAtaactccaggagTATATAAACTCAT
2	YPR133w-A/Tom5:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAtaactccaggagTATATAAACTCAT
2	YPR133w-A/Tom5:TAvariants_5_5_	ggttctggatcaggttcaAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAAAGCAATGGAAAtaactccaggagTATATAAACTCAT
2	YPR133w-A/Tom5:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAAAGCAATGGAAAtaactccaggagTATATAAACTCAT
2	YPR183w/Dpm1:TAvariants_1_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATTTGGTCTTTtaactccaggagTATATAAACTCAT
2	YPR183w/Dpm1:TAvariants_5_1_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATtaactccaggagTATATAAACTCAT
2	YPR183w/Dpm1:TAvariants_5_5_	ggttctggatcaggttcaTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATTTGGTCTTTtaactccaggagTATATAAACTCAT
2	YPR183w/Dpm1:TAvariants_5_5_ggtggaccaggt	ggttctggatcaggttcaggtggaccaggtTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATTTGGTCTTTtaactccaggagTATATAAACTCAT
3	YAL014c/Syn8:length_22_V	ggttctggatcaggttcaAGGTTCAAGGATAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTAGTTGTCGTATTAtaactccaggagTATATAAACTCAT
3	YAL014c/Syn8:length_24_V	ggttctggatcaggttcaAGGTTCAAGGATAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTAGTTGTCGTAGTTGTCTTAtaactccaggagTATATAAACTCAT

3	YAL014c/Syn8:length_26_V	ggttctggatcaggttcaAGGTTCAAGGATAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTAGTTGTCGTAGT TGTCGTAGTTTTAtaactccaggagTATATAAACTCAT
3	YAL014c/Syn8:length_26_V_shuffled_1	ggttctggatcaggttcaAGGTTCAAGGATAACGTACTTCTACTGATTGTGAACGTGTTAGTTGTCGTCGTTGGCGTACTCTTGGTAATACTTTGCATTGT GGTTGTGCTGTTAtaactccaggagTATATAAACTCAT
3	YAL014c/Syn8:length_26_V_shuffled_2	ggttctggatcaggttcaAGGTTCAAGGATAACCTTCTTATTAACTTAGTCGTAGGCGTGCTGGTTGTCTGCGTGGTGATACTAGTAGTACTCGTTGTGA TTGTTCTGTTGTTAtaactccaggagTATATAAACTCAT
3	YAL014c/Syn8:length_26_VA	ggttctggatcaggttcaAGGTTCAAGGATAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTAGTTGCTGTCGC AGTAGCCGTTTTAtaactccaggagTATATAAACTCAT
3	YAL014c/Syn8:TMD_shuffled_1	ggttctggatcaggttcaAGGTTCAAGGATAACCTTATTGTGATACTTTGCATTTTACTGCTCGTGGGCGTGGTGGTATTGCTGCTAAACTTAtaactccagg agTATATAAACTCAT
3	YAL014c/Syn8:TMD_shuffled_2	ggttctggatcaggttcaAGGTTCAAGGATAACGTACTTATTTTGGTGCTGCTGATAAACGTGTTACTTGTGTGCCTAGTGATTCTCGGCTTAtaactccagg agTATATAAACTCAT
3	YAL030w/Snc1:length_24_V	ggttctggatcaggttcaCTAAAAATGAAGATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTGTT GTCCGAtaactccaggagTATATAAACTCAT
3	YAL030w/Snc1:length_26_V	ggttctggatcaggttcaCTAAAAATGAAGATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTGTT GTCGTAGTTCGAtaactccaggagTATATAAACTCAT
3	YAL030w/Snc1:length_26_V_shuffled_1	ggttctggatcaggttcaCTAAAAATGAAGATGGCTGTCGTTATCATTTGTTTTCCCGTTTTAAGTCTTGTCGTAATAATCGTACACGTTATCGTAGCTGTT TTGATCCTGCGAtaactccaggagTATATAAACTCAT
3	YAL030w/Snc1:length_26_V_shuffled_2	ggttctggatcaggttcaCTAAAAATGAAGATGTTAGTTGCTATAGTTGTCCTTGTAGTCGTTTGTATCAGTGTAATTATCGCTATCATCGTAGTTTTTTTG CTGCCCCACCGAtaactccaggagTATATAAACTCAT
3	YAL030w/Snc1:length_26_VA	ggttctggatcaggttcaCTAAAAATGAAGATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTGTT GCTGTCGCACGAtaactccaggagTATATAAACTCAT
3	YAL030w/Snc1:TMD_shuffled_1	ggttctggatcaggttcaCTAAAAATGAAGATGTTAGTTCTGTTGAGTCTTCCCTTTATTGCTATCATCGTAATCGTTATAGCTATCCACGTCGTATGTCGA taactccaggagTATATAAACTCAT
3	YAL030w/Snc1:TMD_shuffled_2	ggttctggatcaggttcaCTAAAAATGAAGATGATCGCTGCTAGTATCGTTGTATGTTTGGTCATAGTATTTATTATCCTGCCCCACTTAATCCTTGTTCGA taactccaggagTATATAAACTCAT
3	YBL091c-A/Scs22:length_22_V	ggttctggatcaggttcaGGTAACGGGCAAAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACGTTGTCT ACtaactccaggagTATATAAACTCAT
3	YBL091c-A/Scs22:length_24_V	ggttctggatcaggttcaGGTAACGGGCAAAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACGTTGTCG TAGTTTACtaactccaggagTATATAAACTCAT
3	YBL091c-A/Scs22:length_26_V	ggttctggatcaggttcaGGTAACGGGCAAAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACGTTGTCG TAGTTGTCGTATACtaactccaggagTATATAAACTCAT
3	YBL091c-A/Scs22:length_26_V_shuffled_1	ggttctggatcaggttcaGGTAACGGGCAAAGTATCGTAATCGTTCTGCTTTGGATCTCCGCAACCGTAGCAGTTTTGTACATAAGTAGAGTCGGCCTCG TCGTCTTGGTTTACtaactccaggagTATATAAACTCAT
3	YBL091c-A/Scs22:length_26_V_shuffled_2	ggttctggatcaggttcaGGTAACGGGCAAAGTTACGTAATCATCGTTGTTGTCGTCAGTACCTCCATCGCAATAGTTTGGCTGCTTCTCTTGGTCGTATT GAGAGGCGCATACtaactccaggagTATATAAACTCAT
3	YBL091c-A/Scs22:length_26_VA	ggttctggatcaggttcaGGTAACGGGCAAAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACGTTGCTG TCGCAGTAGCCTACtaactccaggagTATATAAACTCAT
3	YBL091c-A/Scs22:TMD_shuffled_1	ggttctggatcaggttcaGGTAACGGGCAAAGTTCCTGGAGAGCAACCCTGATCGCAAGTTACATCTTGTTGATAGTCCTTCTCATCGTTGGCTACtaactc caggagTATATAAACTCAT
3	YBL091c-A/Scs22:TMD_shuffled_2	ggttctggatcaggttcaGGTAACGGGCAAAGTTTGTTGTCCATCATAGCAATCCTGACCATCGTTTGGCTCGGCCTTAGAAGTGCAGTCTACTACtaactc caggagTATATAAACTCAT
3	YBL100c:length_22_V	ggttctggatcaggttcaGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTGTTGTCGTAGTTGTCTG Ttaactccaggag TATATAAACTCAT
3	YBL100c:length_24_V	ggttctggatcaggttcaGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTGTTGTCGTAGTTGTCGT AGTTTGTtaactccaggagTATATAAACTCAT
3	YBL100c:length_26_V	ggttctggatcaggttcaGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTGTTGTCGTAGTTGTCGT AGTTGTCGTATGTtaactccaggagTATATAAACTCAT
3	YBL100c:length_26_V_shuffled_1	ggttctggatcaggttcaGATAGAGAACGAATAGTATTTGCAGTCTGTTTATTTGTCGTTGTTGCAATATGTCGAGTCGTAGTAGCGGTTGCCTTTTATAT TATTGTAAACTGTtaactccaggagTATATAAACTCAT
3	YBL100c:length_26_V_shuffled_2	ggttctggatcaggttcaGATAGAGAACGAATAATTTTTTATTGTTTAGTAGTAGTAGTCGCACGAGCGGTCGCAATAGTTTGTGTTAACGTTTTTTTTGT CGCCGTAATTTGTtaactccaggagTATATAAACTCAT
3	YBL100c:length_26_VA	ggttctggatcaggttcaGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTGTTGCTGTCGCAGTAGC CGTTGCTGTCTGTtaactccaggagTATATAAACTCAT
3	YBL100c:TMD_shuffled_1	ggttctggatcaggttcaGATAGAGAACGAATATTTAACGTAGCATATGCGTGTATTGCCTTTCGATTTTTATGTATAGCAATTTGTTACTGTTCTTGCtaa ctccaggagTATATAAACTCAT
3	YBL100c:TMD_shuffled_2	ggttctggatcaggttcaGATAGAGAACGAATATTAATATTTGTATGTATTCGATTTTATGCCGCAAACTTTGCAATTTGTGCGTGTTACTGTTCTTGCtaa ctccaggagTATATAAACTCAT
3	YBR016w:all_C_to_A	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTGCACTGGCTGCAGCCCTGGCTGCATTAGCTATAGCAGCCACCATGGATATGCTATTCtaactcca ggagTATATAAACTCAT
3	YBR016w:all_C_to_S	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCACTGGCTGCATCCCTGGCTGCATTATCTATATCATCCACCATGGATATGCTATTCtaactccagg agTATATAAACTCAT
3	YBR016w:all_C_to_V	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTGTCCTGGCTGCAGTACTGGCTGCATTAGTTATAGTCGTAACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:C1_to_S	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCCCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:C2_to_S	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATCTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:C3_to_S	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATCCATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:C4_to_S	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATCATGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:C5_to_S	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTCCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos1	ggttctggatcaggttcaCAGAGGGGTAACGAATTGTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos10	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos11	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATTGATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos12	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTTTATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos13	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATTGTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos14	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTTAACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos15	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCTTGATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos16	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCTTAGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos17	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGTTGATGCTATTCtaactccag gagTATATAAACTCAT

3	YBR016w:L_pos18	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATTTACTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos19	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGTTGTTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos2	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTTACTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos3	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTTTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos4	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGTTAGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos5	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTTTGTGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos6	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATTACTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos7	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTTTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos8	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGTTAGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:L_pos9	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTTTGTTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:length_22_V	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTAGTTGTCGTAT TCtaactccaggagTATATAAACTCAT
3	YBR016w:length_24_V	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTAGTTGTCGTA GTTGTCTTCtaactccaggagTATATAAACTCAT
3	YBR016w:length_26_V	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTAGTTGTCGTA GTTGTCGTAGTTTTCtaactccaggagTATATAAACTCAT
3	YBR016w:length_26_V_shuffled_1	ggttctggatcaggttcaCAGAGGGGTAACGAAATGTGTGTTGGTCTGGCTGCAGTCGCACTATTATGTGTTATAGTTATGACCTGCTGCGTAGTAGTCG ATTGTCTGGCTTTCtaactccaggagTATATAAACTCAT
3	YBR016w:length_26_V_shuffled_2	ggttctggatcaggttcaCAGAGGGGTAACGAAGATGTCGCTACCGCAGTCTGTCTATGTATGTGCATGGGTTGCGCAGCTGTTCTGTTATGTGTTCTGG TAGTAGTTATATTCtaactccaggagTATATAAACTCAT
3	YBR016w:length_26_VA	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTAGTTGCTGTCG CAGTAGCCGTTTTCtaactccaggagTATATAAACTCAT
3	YBR016w:TMD_shuffled_1	ggttctggatcaggttcaCAGAGGGGTAACGAAGCATTAATGCTGTGTTGCATAGCAGCTCTGATGGATTGCTGTGCTACCGGTCTATGTTTCtaactccag gagTATATAAACTCAT
3	YBR016w:TMD_shuffled_2	ggttctggatcaggttcaCAGAGGGGTAACGAACTGTGCATGGGTGCTTGTTGCTTAGATCTATGTGCTGCAACCGCAATGTGTCTGATATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos1	ggttctggatcaggttcaCAGAGGGGTAACGAAGTCTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos10	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCAGTCTGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos11	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTAGTAATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos12	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTGTTTGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos13	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATAGTCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos14	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCGTAACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos15	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCGTTATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos16	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCGTCGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos17	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGTAATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos18	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATGTTCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos19	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGGTCTTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos2	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTGTACTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos3	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTGTTGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos4	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGTCGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos5	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGTATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos6	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCAGTTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos7	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTGTCGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos8	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGTAGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YBR016w:V_pos9	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGTTTTATGTATATGCTGCACCATGGATATGCTATTCtaactccag gagTATATAAACTCAT
3	YbR056W-A/Mnc1:all_C_to_A	ggttctggatcaggttcaCCACCAAGAAATGATGCAGCCGCTGCAGCCAACGCTGGCGATGCAGCCTCAGCAATTGCGAACGTTTTAGCTGCACTGGCC CTAATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:all_C_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTCATCCTCTTCATCCAACTCTGGCGATTCATCCTCAGCAATTGCGAACGTTTTATCTTCACTGTCCCTA ATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:all_C_to_V	ggttctggatcaggttcaCCACCAAGAAATGATGTCGTAGTTGTCGTAAACGTTGGCGATGTCGTATCAGCAATTGCGAACGTTTTAGTTGTCCTGGTAC TAATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C1_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTCATGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C10_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTCCCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C11_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTCACT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C2_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTCCTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C3_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTCTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C4_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTCATGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT

3	YbR056W-A/Mnc1:C5_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTCCAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C6_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTCAGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C7_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTCATGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C8_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTCCTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:C9_to_S	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATCATGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos1	ggttctggatcaggttcaCCACCAAGAAATGATTTGTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos10	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTTATGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos11	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTTGTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos12	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTTAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos13	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCATTGATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos14	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCATTAGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos15	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTTTGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos16	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGTTAGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos17	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACTTGTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos18	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos19	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATTGTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos2	ggttctggatcaggttcaCCACCAAGAAATGATTGTTTATGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos20	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTTACTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos21	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTTTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos22	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTTACT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos23	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTTT GATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos3	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTTGTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos4	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTTATGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos5	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTTGAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos6	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTTATGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos7	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTTGGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos8	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTTTAGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCTA ATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:L_pos9	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCTTGTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:length_24_V	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AGTTATTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:length_26_V	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AGTTGTCGTAATTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:length_26_V_shuffled_1	ggttctggatcaggttcaCCACCAAGAAATGATTGTGCAATTTGTCTGTGTTGTTGTGCGGTCTGTCTAGGCTGTGTTTGTTGCGATTGTGTTGTAAACTG TTTATCAAACATTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:length_26_V_shuffled_2	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTGTCGATTGTGGCAACGTTGTATTATCATGTTGTTGTTGTATTAACGCGTGTTGCGTTTGTTG TCTGGCACTAATTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:length_26_VA	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AGTTGCTGTCATTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:TMD_shuffled_1	ggttctggatcaggttcaCCACCAAGAAATGATAACTGTTGCCTATGTTGTTTAGCGTGTTGTTGTAACTGTATTTGTTCACTGGGCGTTGCATGTGATTG TATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:TMD_shuffled_2	ggttctggatcaggttcaCCACCAAGAAATGATCTAAACGGCCTGTGTTTATGTTGTTGTTGTGCATCAGTTAACTGTTGCTGTATTTGTGCGTGTTGTGA TATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos1	ggttctggatcaggttcaCCACCAAGAAATGATGTCTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos10	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATGTCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos11	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCGTATCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos12	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTGTTGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos13	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGTCATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos14	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAGTAGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos15	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGTTAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos16	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGGTCGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos17	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTATTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos18	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTGTTTGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos19	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTAGTCTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT

3	YbR056W-A/Mnc1:V_pos2	ggttctggatcaggttcaCCACCAAGAAATGATTGTGTATGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos20	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTGTACTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos21	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTGTTTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos22	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGGTCCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos23	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTGT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos3	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTGTTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos4	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTGTCTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos5	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTGTAAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos6	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTGTTTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos7	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACGTCGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos8	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGTAGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YbR056W-A/Mnc1:V_pos9	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGTTTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
3	YBR162w-A/Ysy6:length_22_V	ggttctggatcaggttcaGTGATATCCAAAACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCGTTCTA taactccaggagTATATAAACTCAT
3	YBR162w-A/Ysy6:length_24_V	ggttctggatcaggttcaGTGATATCCAAAACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCGTTGTC GTACTAtaactccaggagTATATAAACTCAT
3	YBR162w-A/Ysy6:length_26_V	ggttctggatcaggttcaGTGATATCCAAAACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCGTTGTC GTAGTTGTCCTAtaactccaggagTATATAAACTCAT
3	YBR162w-A/Ysy6:length_26_V_shuffled_1	ggttctggatcaggttcaGTGATATCCAAAACTCAAGTAGGTGTTCTTGTCCTGATCGTCATCTATCTATTTCTCCTTGTTTGGTTGGGTTTGGGTAGCGTA GTTGGTATTCTAtaactccaggagTATATAAACTCAT
3	YBR162w-A/Ysy6:length_26_V_shuffled_2	ggttctggatcaggttcaGTGATATCCAAAACTGTTTTGGGTAGCCTGGGTTTGCTAATCTTTATCCAAGTTGTAGGTATTCTCTATGTTGGTGTACTTGTC GTCCTTTGGCTAtaactccaggagTATATAAACTCAT
3	YBR162w-A/Ysy6:length_26_VA	ggttctggatcaggttcaGTGATATCCAAAACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCGTTGCT GTCGCAGTACTAtaactccaggagTATATAAACTCAT
3	YBR162w-A/Ysy6:TMD_shuffled_1	ggttctggatcaggttcaGTGATATCCAAAACTCTTTGGCAATTTGTTCTTAGCTTGATTGGTGGTGGTATCATCTTGCTGCTCTATCTAGGTGTACTAtaac tccaggagTATATAAACTCAT
3	YBR162w-A/Ysy6:TMD_shuffled_2	ggttctggatcaggttcaGTGATATCCAAAACTTTGGGTGTACTGCTAGGTTTGTATGGTTTTATCGTTATCCTCTGGCTTATTAGCCTTGGTCAACTAtaac tccaggagTATATAAACTCAT
3	YDL012c:all_C_to_A	ggttctggatcaggttcaAGTTCTGGAAATGAAGACGCACTAGCGGGCGCCCTAGCGGGTTTAGCTTTAGCAGCCACTTTGGATATGCTGTTTtaactccag gagTATATAAACTCAT
3	YDL012c:all_C_to_S	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTCACTAGCGGGCTCCCTAGCGGGTTTATCTTTATCATCCACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:all_C_to_V	ggttctggatcaggttcaAGTTCTGGAAATGAAGACGTCCTAGCGGGCGTACTAGCGGGTTTAGTTTTAGTCGTAACTTTGGATATGCTGTTTtaactccag gagTATATAAACTCAT
3	YDL012c:C1_to_S	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTCCCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:C2_to_S	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTCTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:C3_to_S	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATCCTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:C4_to_S	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATCATGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:C5_to_S	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTCCACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos1	ggttctggatcaggttcaAGTTCTGGAAATGAATTGTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos10	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos11	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATTGTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos13	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATTGTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos14	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTTAACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos15	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTTTGTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos16	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTAGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos17	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGTTGATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos18	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATTTACTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos19	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGTTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos2	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTTACTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos20	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTAtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos3	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTTTGGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos4	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTATTAGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos5	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGTTGTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos6	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTTACTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos7	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTTTGGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:L_pos8	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTATTAGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT

3	YDL012c:L_pos9	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGTTGTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:length_22_V	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTGTTGTCta actccaggagTATATAAACTCAT
3	YDL012c:length_24_V	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTGTTGTCGT AGTTtaactccaggagTATATAAACTCAT
3	YDL012c:length_26_V	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTGTTGTCGT AGTTGTCGTAtaactccaggagTATATAAACTCAT
3	YDL012c:length_26_V_shuffled_1	ggttctggatcaggttcaAGTTCTGGAAATGAAGTTGCGTTTCTAATGGACTTAGGTGTTTGTGCGGATCTGTTGCTATGTACTGTAGTCTTAGTCTGCTG TTGTGTAGGCtaactccaggagTATATAAACTCAT
3	YDL012c:length_26_V_shuffled_2	ggttctggatcaggttcaAGTTCTGGAAATGAACTATTATGTTGTACTTGCGCGGTATGTTGTTTAGATGACGTTCTAATGGTCGGCTTTGTCCTGGGTGT TGCGGTATTGtaactccaggagTATATAAACTCAT
3	YDL012c:length_26_VA	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTGTTGCTGT CGCAGTAGCCtaactccaggagTATATAAACTCAT
3	YDL012c:TMD_shuffled_1	ggttctggatcaggttcaAGTTCTGGAAATGAAGCGTGTTGTCTAATGCTGTGCTTATTATTTGACGGTGCGTTGACTGATTGTCTATGTGGCtaactccagg agTATATAAACTCAT
3	YDL012c:TMD_shuffled_2	ggttctggatcaggttcaAGTTCTGGAAATGAAGCGGATTGTTTACTGCTAGACATGTTAGCGGGTTGCCTATTGTGTTTTGGCACTTGTTGTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos1	ggttctggatcaggttcaAGTTCTGGAAATGAAGTCTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos10	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTGTCTGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos11	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTAGTATTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos12	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTGTTTGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos13	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTAGTCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos14	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCGTAACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos15	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTGTTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos16	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTGTCGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos17	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGTAATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos18	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATGTTCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos19	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGGTCTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos2	ggttctggatcaggttcaAGTTCTGGAAATGAAGACGTACTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos20	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGGTAtaactccag gagTATATAAACTCAT
3	YDL012c:V_pos3	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTGTTGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos4	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGTCGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos5	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGTATGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos6	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCGTTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos7	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTGTCGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos8	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGTAGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL012c:V_pos9	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGTTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccagg agTATATAAACTCAT
3	YDL241w:length_22_V	ggttctggatcaggttcaAGGCCTTCTCATTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCGTTGTCGTAGTTGTCGTAtaac tccaggagTATATAAACTCAT
3	YDL241w:length_24_V	ggttctggatcaggttcaAGGCCTTCTCATTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCGTTGTCGTAGTTGTCGTAGTT GTCtaactccaggagTATATAAACTCAT
3	YDL241w:length_26_V	ggttctggatcaggttcaAGGCCTTCTCATTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCGTTGTCGTAGTTGTCGTAGTT GTCGTAGTTtaactccaggagTATATAAACTCAT
3	YDL241w:length_26_V_shuffled_1	ggttctggatcaggttcaAGGCCTTCTCATTCTTATGTTTTCGTCACCGTGTTGCTCGTCGTAGTCGCCCTGGTAGTCATTGTAATATTTACAGTTATGGTT GTTCTAAAAtaactccaggagTATATAAACTCAT
3	YDL241w:length_26_V_shuffled_2	ggttctggatcaggttcaAGGCCTTCTCATTCTGTTGTCGTCTATAAATTCATGCTAGTATTTATTGCCGTTGTCGTACTGACATTGCTCGTCGTTATAGTT GTAGTGACCtaactccaggagTATATAAACTCAT
3	YDL241w:length_26_VA	ggttctggatcaggttcaAGGCCTTCTCATTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCGTTGCTGTCGCAGTAGCCGTT GCTGTCGCAtaactccaggagTATATAAACTCAT
3	YDR034W-B:all_C_to_A	ggttctggatcaggttcaAGAGAATCCGGTGGTGCAGCCAGAACCGCTGCACACTTCCTAGCCGCTCTAGCATTAATTAACCTAGCCGCTGACGTTTTTta actccaggagTATATAAACTCAT
3	YDR034W-B:all_C_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTCATCCAGAACCTCTTCACACTTCCTATCCTCTCTATCATTAATTAACCTATCCTCTGACGTTTTTtaact ccaggagTATATAAACTCAT
3	YDR034W-B:all_C_to_V	ggttctggatcaggttcaAGAGAATCCGGTGGTGTCGTAAGAACCGTTGTCCACTTCCTAGTAGTTCTAGTCTTAATTAACCTAGTAGTTGACGTTTTTtaa ctccaggagTATATAAACTCAT
3	YDR034W-B:C1_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTCATGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:C2_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTCCAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:C3_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTCCTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:C4_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTCTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:C5_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATCATGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:C6_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTCCCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:C7_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATCATTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:C8_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATCTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT

3	YDR034W-B:C9_to_S	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTCAGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos1	ggttctggatcaggttcaAGAGAATCCGGTGGTTTGTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos10	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATTATGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos11	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTTGCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos12	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTTTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos13	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATTGTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos14	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos15	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTATTGAACCTATGTTGTGACGTTTTTtaa ctccaggagTATATAAACTCAT
3	YDR034W-B:L_pos16	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTTTACTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos17	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACTTGTGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos18	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATTATGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos19	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTTGGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos2	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTTAAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos3	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTTTGACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos4	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGATTATGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos5	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTTGTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos6	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTTACACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos7	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTTTGTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos8	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTACTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:L_pos9	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCTTGTGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:length_22_V	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGTTGTCGTAGA CtaactccaggagTATATAAACTCAT
3	YDR034W-B:length_24_V	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGTTGTCGTAGTT GTCGACtaactccaggagTATATAAACTCAT
3	YDR034W-B:length_26_V	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGTTGTCGTAGTT GTCGTAGTTGACtaactccaggagTATATAAACTCAT
3	YDR034W-B:length_26_V_shuffled_1	ggttctggatcaggttcaAGAGAATCCGGTGGTAGACTATGTGTTTGTATTTGTTGCCACTGTGTTGTACTATTCGTTTTAGTATGTGTCTGTAACTGTGT CCTAACCTGTGACtaactccaggagTATATAAACTCAT
3	YDR034W-B:length_26_V_shuffled_2	ggttctggatcaggttcaAGAGAATCCGGTGGTCTATTCATTTGTGTTGTAGTTTGTGTCTGTCTATGTCACTGTTGTTGCCTATTAACCGTATGTGTCGTT AACAGATGTGACtaactccaggagTATATAAACTCAT
3	YDR034W-B:length_26_VA	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGTTGCTGTCGC AGTAGCCGTTGACtaactccaggagTATATAAACTCAT
3	YDR034W-B:TMD_shuffled_1	ggttctggatcaggttcaAGAGAATCCGGTGGTTTATGTATTTGTTGTAACCTACTATGTCTATGCTGTTGTTGTTGTTTCAGAACCCACGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:TMD_shuffled_2	ggttctggatcaggttcaAGAGAATCCGGTGGTAACTGTAGATGTTTCTGTCTATGTTGTCACACCTGTTGTTGTATTTTACTATGCCTAGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos1	ggttctggatcaggttcaAGAGAATCCGGTGGTGTCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos10	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTAGTCTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos11	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTGTACTATGTTTAATTAACCTATGTTGTGACGTTTTTtaa ctccaggagTATATAAACTCAT
3	YDR034W-B:V_pos12	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTGTTTGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos13	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTAGTCTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos14	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTGTAATTAACCTATGTTGTGACGTTTTTtaa ctccaggagTATATAAACTCAT
3	YDR034W-B:V_pos15	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAGTTAACCTATGTTGTGACGTTTTTtaa ctccaggagTATATAAACTCAT
3	YDR034W-B:V_pos16	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTGTCCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos17	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACGTATGTTGTGACGTTTTTtaa ctccaggagTATATAAACTCAT
3	YDR034W-B:V_pos18	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTAGTTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos19	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTGTCGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos2	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCGTAAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaa ctccaggagTATATAAACTCAT
3	YDR034W-B:V_pos3	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTGTTACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos4	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAGTCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos5	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCGTATGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaa ctccaggagTATATAAACTCAT
3	YDR034W-B:V_pos6	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTGTTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos7	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTGTCTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
3	YDR034W-B:V_pos8	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACGTACTATGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaa ctccaggagTATATAAACTCAT

3	YDR034W-B:V_pos9	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCGTTTGTTGTCTATGTTTAATTAACCTATGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT
3	YDR086c/Sss1:length_22_V	ggttctggatcaggttcaAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCGTTGTCGTAAAGtaactccaggagTATATAAACTCAT
3	YDR086c/Sss1:length_24_V	ggttctggatcaggttcaAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCGTTGTCGTAGTTGTCAAGtaactccaggagTATATAAACTCAT
3	YDR086c/Sss1:length_26_V	ggttctggatcaggttcaAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCGTTGTCGTAGTTGTCGTAGTTAAGtaactccaggagTATATAAACTCAT
3	YDR086c/Sss1:length_26_V_shuffled_1	ggttctggatcaggttcaAAGGAATACACCAAGGCCTTTGCTGTTGTAGTTGTAGTTGTCATTATTGGTGGTGTCGGTGGTATTGTTATCATTGTCTACATCGCAAAGGTCAAGtaactccaggagTATATAAACTCAT
3	YDR086c/Sss1:length_26_V_shuffled_2	ggttctggatcaggttcaAAGGAATACACCAAGATTATTAAGGTCGTAGGTGCTGTTATCGGTGGTGCCGGTGTCGTTGTCATTGTAGTTGTTGCAGTCTACATCTTTATTAAGtaactccaggagTATATAAACTCAT
3	YDR086c/Sss1:length_26_VA	ggttctggatcaggttcaAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCGTTGCTGTCGCAGTAGCCGTTAAGtaactccaggagTATATAAACTCAT
3	YDR086c/Sss1:TMD_shuffled_1	ggttctggatcaggttcaAAGGAATACACCAAGTACGTCATTAAGGCTATTATTGCCGCAGGTTTTGTTGGTATTATCATCGGTGGTGTCAAGTTGATTCATATTtaactccaggagTATATAAACTCAT
3	YDR086c/Sss1:TMD_shuffled_2	ggttctggatcaggttcaAAGGAATACACCAAGGCAGGTATCGCCAAGTACGTCATTATCATTGGTATTATTGTTGGTGTCGCTTTTGGTAAGTTGATTCATATTtaactccaggagTATATAAACTCAT
3	YDR200c/Far9:length_22_V	ggttctggatcaggttcaACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTGTTGTCGTAGTTAAGtaactccaggagTATATAAACTCAT
3	YDR200c/Far9:length_24_V	ggttctggatcaggttcaACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTGTTGTCGTAGTTGTCGTAAAGtaactccaggagTATATAAACTCAT
3	YDR200c/Far9:length_26_V	ggttctggatcaggttcaACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTGTTGTCGTAGTTGTCGTAGTTGTCAAGtaactccaggagTATATAAACTCAT
3	YDR200c/Far9:length_26_V_shuffled_1	ggttctggatcaggttcaACAAGAGTTAGTAAGGGTGTTGGAGTTGGAATGTTAGCCTTTGTATTCGTCGTTGTTGCCGTAATTCTCGCGGTCTCAGTAACGGTCGTTGTTAAGtaactccaggagTATATAAACTCAT
3	YDR200c/Far9:length_26_V_shuffled_2	ggttctggatcaggttcaACAAGAGTTAGTAAGATTACGGTACTCGTTTTCATGGTCGGTGTTTTTGTTTTAGTATCAGCCGCCGTCGTTGGAGTCGTTGGAGTAGTTGCGAAGtaactccaggagTATATAAACTCAT
3	YDR200c/Far9:length_26_VA	ggttctggatcaggttcaACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTGTTGCTGTCGCAGTAGCCGTTGCTAAGtaactccaggagTATATAAACTCAT
3	YDR200c/Far9:TMD_shuffled_1	ggttctggatcaggttcaACAAGAGTTAGTAAGGCCTCAGCGCTCATTGCCTTTGGAGGTGTTTTATTCGGAGTTGTAGTTATGACGAAGCAATTGCCACAAtaactccaggagTATATAAACTCAT
3	YDR200c/Far9:TMD_shuffled_2	ggttctggatcaggttcaACAAGAGTTAGTAAGTTCTCAGGTGGAGTAATTTTAATGACGGTTGGATTTGCGGTTCTCGCCGTTGCCAAGCAATTGCCACAAtaactccaggagTATATAAACTCAT
3	YDR210w:all_C_to_A	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCGCATTGGACAGTGCCTTAAAAGCTTTGGCAGCCGCTTTCTTATTAGAATTGGTATGCGATTaactccaggagTATATAAACTCAT
3	YDR210w:all_C_to_S	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTCATTGGACAGTTCCTTAAAATCTTTGTCATCCTCTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:all_C_to_V	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCGTCTTGGACAGTGTATTAAAAGTTTTGGTCGTAGTTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:C1_to_S	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTCATTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:C2_to_S	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTCCTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:C3_to_S	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATCCTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:C4_to_S	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTCATGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:C5_to_S	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTCCTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:C6_to_S	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTCTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos1	ggttctggatcaggttcaCAAGGTCAACCAAAGTTGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos10	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTATTATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos11	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATTGTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos12	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTATGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos13	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTTGTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos14	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTTATGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos15	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTTGTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos16	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTATTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos17	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTGTTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos2	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGTTAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos3	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAATTGTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos4	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTTATTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos5	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos6	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGTTAAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos7	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACTTGTGTTTAAAATGTTTGTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos8	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTTATTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:L_pos9	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTGAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCGATtaactccaggagTATATAAACTCAT
3	YDR210w:length_22_V	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGTTGTCGTAGTTGAAtaactccaggagTATATAAACTCAT
3	YDR210w:length_24_V	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGTTGTCGTAGTTGTCGTAGAAtaactccaggagTATATAAACTCAT
3	YDR210w:length_26_V	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGTTGTCGTAGTTGTCGTAGTTGTCGAAtaactccaggagTATATAAACTCAT

3	YDR210w:length_26_V_shuffled_1	ggttctggatcaggttcaCAAGGTCAACCAAAGGTCGTAGTCGAGGACTGTTGTGTCGTTTTATTATGTTTGTTCAGTTGTTTGGTTAAAGTTAGCGTATG TTTATGCGAAGAAtaactccaggagTATATAAACTCAT
3	YDR210w:length_26_V_shuffled_2	ggttctggatcaggttcaCAAGGTCAACCAAAGTTAGTTTGTAAAGACGTAAGCTGTGAATGCTGTGAGGTCAGTGTCTGTTTATTGTGTGTAGTTGTTG TCTTGTTCTTAGAAtaactccaggagTATATAAACTCAT
3	YDR210w:length_26_VA	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGTTGCTGTCGCAG TAGCCGTTGCTGAAtaactccaggagTATATAAACTCAT
3	YDR210w:TMD_shuffled_1	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGACTGTTTGGAATGTAGTAGCTGCTTGTTCTGTTTATGTTTATTATGTAAAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:TMD_shuffled_2	ggttctggatcaggttcaCAAGGTCAACCAAAGGAATGTAAATTGTGTTTATTCTGTTGTTTAGACTGCTTGAGTTTATGTAGCGAGGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos1	ggttctggatcaggttcaCAAGGTCAACCAAAGGTCGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos10	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAGTCTGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos11	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAAGTATTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos12	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTGTTTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos13	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGGTCTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos14	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTGTATGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos15	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTGTTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos16	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTGTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos17	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTTCGTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos18	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTAGTTGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos2	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGTAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos3	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAGTTTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos4	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCGTCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos5	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCGTAGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos6	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGTTAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos7	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACGTCTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos8	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTGTATTAAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR210w:V_pos9	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTGTTAAATGTTTGTGTTGTTGTTTCTTATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
3	YDR281c/Phm6:length_24_V	ggttctggatcaggttcaGGGTTTCTCGAGGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGGTT GTCATGtaactccaggagTATATAAACTCAT
3	YDR281c/Phm6:length_26_V	ggttctggatcaggttcaGGGTTTCTCGAGGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGGTT GTCGTAGTTATGtaactccaggagTATATAAACTCAT
3	YDR281c/Phm6:length_26_V_shuffled_1	ggttctggatcaggttcaGGGTTTCTCGAGGAATTGACACTCTTGCTTATTTTGAGTGTAATCTATCTGGTAGTGGTGGTCGGTTACGTTATCGTCATGAT TTTTATTGTTATGtaactccaggagTATATAAACTCAT
3	YDR281c/Phm6:length_26_V_shuffled_2	ggttctggatcaggttcaGGGTTTCTCGAGGAATTGCTCTATACAATTTACGTGATTATGTTGGTTGTCATCTTTAGTGTGGGTCTTTTGGTAGTTATTGTA GTCATCCTGATGtaactccaggagTATATAAACTCAT
3	YDR281c/Phm6:length_26_VA	ggttctggatcaggttcaGGGTTTCTCGAGGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGGTT GCTGTCGCAATGtaactccaggagTATATAAACTCAT
3	YDR281c/Phm6:TMD_shuffled_1	ggttctggatcaggttcaGGGTTTCTCGAGGAATTGATCTATGTAATTCTCTTGGTCCTGATGTTTATTCTTGTGTTGACAATTAGTATCTACGGTGTGAT GACGATGACGAAGtaactccaggagTATATAAACTCAT
3	YDR281c/Phm6:TMD_shuffled_2	ggttctggatcaggttcaGGGTTTCTCGAGGAACTCACAGGTTTGAGTATCGTAGTCATGATCTTGGTGTTGGTGCTGTATTACTTTATTATTCTTATTAT GACGATGACGAAGtaactccaggagTATATAAACTCAT
3	YDR468c/Tlg1:length_22_V	ggttctggatcaggttcaGAAAAATACGACGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGTTGTCGTAGCT taactccaggagTATATAAACTCAT
3	YDR468c/Tlg1:length_24_V	ggttctggatcaggttcaGAAAAATACGACGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGTTGTCGTAGTT GTCGCTtaactccaggagTATATAAACTCAT
3	YDR468c/Tlg1:length_26_V	ggttctggatcaggttcaGAAAAATACGACGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGTTGTCGTAGTT GTCGTAGTTGCTtaactccaggagTATATAAACTCAT
3	YDR468c/Tlg1:length_26_V_shuffled_1	ggttctggatcaggttcaGAAAAATACGACGATGTTTTGGTCATAGGAGTTGTAATTTTATGTGTCGTTCTTGTTTTAGTCTGTATAGTCATTTTCGTTGCA CTTTTGGTAGCTtaactccaggagTATATAAACTCAT
3	YDR468c/Tlg1:length_26_V_shuffled_2	ggttctggatcaggttcaGAAAAATACGACGATGTTTTACTTTTGTGTGTCGTCGTAGGAGTTATTGTTGTTATATTGATTGTTATAGTCTTATGTGTCGC ACTTTTCGTAGCTtaactccaggagTATATAAACTCAT
3	YDR468c/Tlg1:length_26_VA	ggttctggatcaggttcaGAAAAATACGACGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGTTGCTGTCGCA GTAGCCGTTGCTtaactccaggagTATATAAACTCAT
3	YDR468c/Tlg1:TMD_shuffled_1	ggttctggatcaggttcaGAAAAATACGACGATTTAGGAATTATAGCATTACTTTTGTTCATACTTTGTTGTGTCGTTGTCATTGTTTTGGCTtaactccagga gTATATAAACTCAT
3	YDR468c/Tlg1:TMD_shuffled_2	ggttctggatcaggttcaGAAAAATACGACGATTTATGTGTCATAGCATTCTGTGTTTTGGGAATAATTTTAGTTCTTTTGGTCCTTATTGCTtaactccagga gTATATAAACTCAT
3	YDR498c/Sec20:length_22_V	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGGTTGTCGTAGTTGTCCG Ttaactccaggag TATATAAACTCAT
3	YDR498c/Sec20:length_24_V	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGGTTGTCGTAGTTGTCGT AGTTCGTtaactccaggagTATATAAACTCAT
3	YDR498c/Sec20:length_26_V	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGGTTGTCGTAGTTGTCGT AGTTGTCGTACGTtaactccaggagTATATAAACTCAT
3	YDR498c/Sec20:length_26_V_shuffled_1	ggttctggatcaggttcaCAAGAGAAACGAGATTGGCTTGTATCAGTCTGCTGCTATTGGCTAGTCCTCGTTGTCGTCGTCTTCGTTGGGCTATTAGTTGT TGTATCGGTACGTtaactccaggagTATATAAACTCAT
3	YDR498c/Sec20:length_26_V_shuffled_2	ggttctggatcaggttcaCAAGAGAAACGAGATTCAGTTTGGCTCGTCGTCTGGGTCCTAGTTGTTGTAGTAGTCCTTTGCTTCCTAGTAGTTTTATGCTA TTCGGTCGGGCGTtaactccaggagTATATAAACTCAT
3	YDR498c/Sec20:length_26_VA	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGGTTGCTGTCGCAGTAG CCGTTGCTGTCCGTtaactccaggagTATATAAACTCAT
3	YDR498c/Sec20:TMD_shuffled_1	ggttctggatcaggttcaCAAGAGAAACGAGATTCGTGCGTCCTACTCTGGTGGGGGTGCCTTTCATTCGTCTTAGTTCTATATCGTCGTATTTTCAAGtaa ctccaggagTATATAAACTCAT
3	YDR498c/Sec20:TMD_shuffled_2	ggttctggatcaggttcaCAAGAGAAACGAGATTTCCTCTGCCTAGTTGTCCTATGCTGGTATTCGTCAGTCTGGCTTGGGTTACGTCGTATTTTCAAGtaa ctccaggagTATATAAACTCAT

3	YER019c-A/Sbh2:length_22_V	ggttctggatcaggttcaTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGGTTAAAtaactccaggagTATATAAACTCAT
3	YER019c-A/Sbh2:length_24_V	ggttctggatcaggttcaTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGGTTGTCGTAAAAtaactccaggagTATATAAACTCAT
3	YER019c-A/Sbh2:length_26_V	ggttctggatcaggttcaTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGGTTGTCGTAGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YER019c-A/Sbh2:length_26_V_shuffled_1	ggttctggatcaggttcaTTCAGAGTCGACTCCTTCGTCCATGTCTTGCTAGTTTCTGTCTTCGGTGTAATTCTTGTCACGGCTTTGCTATCCGTAATCGTGGTTTTGTTCAAAtaactccaggagTATATAAACTCAT
3	YER019c-A/Sbh2:length_26_V_shuffled_2	ggttctggatcaggttcaTTCAGAGTCGACTCCATTATCGTGTTCTTCGGTTTGGTCTTGTTCGTTGTCGTTGTCCTACTACATGCTTCCTCTGTAACGGTCTTGCTTGTAAAAtaactccaggagTATATAAACTCAT
3	YER019c-A/Sbh2:length_26_VA	ggttctggatcaggttcaTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGGTTGCTGTCGCAGTAAAAtaactccaggagTATATAAACTCAT
3	YER019c-A/Sbh2:TMD_shuffled_1	ggttctggatcaggttcaTTCAGAGTCGACTCCACGCATTTCCTACTTTTGATTGTGGTATTCGGTCTATTGATCTTGTCTGCTGTCTTCGTCTCCAAATTTACACACATTtaactccaggagTATATAAACTCAT
3	YER019c-A/Sbh2:TMD_shuffled_2	ggttctggatcaggttcaTTCAGAGTCGACTCCCTATTGTCTCATATTTTGTTGTCCGCTTTCGTCGTATTCATCACGCTACTTTTCGTCGGTGTGAAATTTACACACATTtaactccaggagTATATAAACTCAT
3	YER087c-B/Sbh1:length_22_V	ggttctggatcaggttcaCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTGTTAAAtaactccaggagTATATAAACTCAT
3	YER087c-B/Sbh1:length_24_V	ggttctggatcaggttcaCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTGTTGTCGTAAAAtaactccaggagTATATAAACTCAT
3	YER087c-B/Sbh1:length_26_V	ggttctggatcaggttcaCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTGTTGTCGTAGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YER087c-B/Sbh1:length_26_V_shuffled_1	ggttctggatcaggttcaCTAAGAGTAGATCCCGTTATTGTTGTCTTCGCAATCGTTTTTGTTTTGGTCCATCTATTAGTAGTTGTGTTTTTAGGTGTTTCTGCGTCTGTCAAAtaactccaggagTATATAAACTCAT
3	YER087c-B/Sbh1:length_26_V_shuffled_2	ggttctggatcaggttcaCTAAGAGTAGATCCCGTTTTTATTGTGTTCGTTGTTGCAGTATTTTCTGCGTTAGGTCTATTATTGGTTATCGTTGTTGTCGTCCATGTCTCTAAAtaactccaggagTATATAAACTCAT
3	YER087c-B/Sbh1:length_26_VA	ggttctggatcaggttcaCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTGTTGCTGTCGCAGTAAAAtaactccaggagTATATAAACTCAT
3	YER087c-B/Sbh1:TMD_shuffled_1	ggttctggatcaggttcaCTAAGAGTAGATCCCTTTATCGTTATTCATTCTGGTTTCGTTGCAGTCTTATTGTCTGCGGTTTTAGTTCTATTTGTGAAAGTTGCCGGTAAGtaactccaggagTATATAAACTCAT
3	YER087c-B/Sbh1:TMD_shuffled_2	ggttctggatcaggttcaCTAAGAGTAGATCCCTCTTTATTGGGTGTTATCGTTTTCTTATCTGTTTTTGTTGCGGCACATGTGGTCCTAATTTTTAAAGTTGCCGGTAAGtaactccaggagTATATAAACTCAT
3	YER100w/Ubc6:length_22_V	ggttctggatcaggttcaCCTAATGATAGTTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGGTTGTCGTAGTTAAAtaactccaggagTATATAAACTCAT
3	YER100w/Ubc6:length_24_V	ggttctggatcaggttcaCCTAATGATAGTTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGGTTGTCGTAGTTGTCGTAAAAtaactccaggagTATATAAACTCAT
3	YER100w/Ubc6:length_26_V	ggttctggatcaggttcaCCTAATGATAGTTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGGTTGTCGTAGTTGTCGTAGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YER100w/Ubc6:length_26_V_shuffled_1	ggttctggatcaggttcaCCTAATGATAGTTCTTTGATGGTTATTTTGATGGTTTTTTTTATTGGTGTTGTAATCGTTGTAGTTTATTCAGTCTTTGCTGTCGGCGTCCTTAAAtaactccaggagTATATAAACTCAT
3	YER100w/Ubc6:length_26_V_shuffled_2	ggttctggatcaggttcaCCTAATGATAGTTCTGCTATCTTGTCAGTTGGTTTTGTTCTTGTTGTCGTTTTTGGCGTATTGATGGTCTATATTATGTTTGTCGTTTGTAATTAAAtaactccaggagTATATAAACTCAT
3	YER100w/Ubc6:length_26_VA	ggttctggatcaggttcaCCTAATGATAGTTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGGTTGCTGTCGCAGTAGCCGTTGCTAAAtaactccaggagTATATAAACTCAT
3	YER120w/Scs2:length_22_V	ggttctggatcaggttcaAATGAATCATCCAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACGTTGTCGTAGTTAGAtaactccaggagTATATAAACTCAT
3	YER120w/Scs2:length_24_V	ggttctggatcaggttcaAATGAATCATCCAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACGTTGTCGTAGTTGTCGTAAGAtaactccaggagTATATAAACTCAT
3	YER120w/Scs2:length_26_V	ggttctggatcaggttcaAATGAATCATCCAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACGTTGTCGTAGTTGTCGTAGTTGTCAGAtaactccaggagTATATAAACTCAT
3	YER120w/Scs2:length_26_V_shuffled_1	ggttctggatcaggttcaAATGAATCATCCAGCGGTGTAGTTGTAATCCTTTACTTATTGATACTCGTTGTTGTTGGAGTTTTCTTGATGATAGTCGTCGTCTTCTGGGCAAGAtaactccaggagTATATAAACTCAT
3	YER120w/Scs2:length_26_V_shuffled_2	ggttctggatcaggttcaAATGAATCATCCAGCTTGGTTATAATGTTCATACTCGTTGTCGTATACCTTGGTTTATTGGTCATCTTCGTTTGGGTAGGAGTTGCAGTTGTCAGAtaactccaggagTATATAAACTCAT
3	YER120w/Scs2:length_26_VA	ggttctggatcaggttcaAATGAATCATCCAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACGTTGCTGTCGCAGTAGCCGTTGCTAGAtaactccaggagTATATAAACTCAT
3	YFL046w/Fmp32:length_22_V	ggttctggatcaggttcaTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGGTTGTCAGGtaactccaggagTATATAAACTCAT
3	YFL046w/Fmp32:length_24_V	ggttctggatcaggttcaTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGGTTGTCGTAGTTAGGtaactccaggagTATATAAACTCAT
3	YFL046w/Fmp32:length_26_V	ggttctggatcaggttcaTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGGTTGTCGTAGTTGTCGTAAGGtaactccaggagTATATAAACTCAT
3	YFL046w/Fmp32:length_26_V_shuffled_1	ggttctggatcaggttcaAAGAGTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGGTTGTCGTAGTTGTCGTAAGGtaactccaggagTATATAAACTCAT
3	YFL046w/Fmp32:length_26_V_shuffled_2	ggttctggatcaggttcaAACAACTTGTAGGTTTTGTCACAGTCCTGGTCCAAGTTGTATGGGCAGTATTATATATGTGTGTCATAGGTGTTATGAGGtaactccaggagTATATAAACTCAT
3	YFL046w/Fmp32:length_26_V_shuffled_2	ggttctggatcaggttcaTCTGTCAAGACTCAAACAGGTGCATTTGTCACACAAGTCGTTATGGTTGTCGGTGTCGTACTTTTAGTATGTTATATACTGTGGGCAATGGTAAGGtaactccaggagTATATAAACTCAT
3	YFL046w/Fmp32:length_26_VA	ggttctggatcaggttcaTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGGTTGCTGTCGCAGTAGCCAGGtaactccaggagTATATAAACTCAT
3	YFL046w/Fmp32:TMD_shuffled_1	ggttctggatcaggttcaTCTGTCAAGACTCAATTTATGTATACAGTCGCACTGGTACAAGGTATACTTACAATGGTCTGTTTAGGTTGGGCAAGGTTATTAACAtaactccaggagTATATAAACTCAT
3	YFL046w/Fmp32:TMD_shuffled_2	ggttctggatcaggttcaTCTGTCAAGACTCAAATGACAGCAATACAAACAGCATTTTGTGTCGGTTGGCTGGTCTATATGGGTTTACTTGTAAGGTTATTAACAtaactccaggagTATATAAACTCAT
3	YGL098w/Use1:length_26_V	ggttctggatcaggttcaAAGAGTAAATTGAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCGTTCCGtaactccaggagTATATAAACTCAT
3	YGL098w/Use1:length_26_V_shuffled_1	ggttctggatcaggttcaAAGAGTAAATTGAGTACACTGTTCTTTATTTTCCTCTTATTTCAAGTTGTTGGAATTATCTACGTGATTATAATTTTGTTTTTTTATATGACTCCGtaactccaggagTATATAAACTCAT
3	YGL098w/Use1:length_26_V_shuffled_2	ggttctggatcaggttcaAAGAGTAAATTGAGTTATATTATCGTTTTCTACATGCAATTTATATTTATTGTTTTGCTCTTCATTTTTATTGTGCTGTTTTTAGGAACAACTCCGtaactccaggagTATATAAACTCAT
3	YGL098w/Use1:TMD_shuffled_1	ggttctggatcaggttcaAAGAGTAAATTGAGTTTTATTTTTTTTTTATGGATTCTTGTACCTCATACAATTTTTAACTATTGTGATTATTATGCTGGTTTTCACAATCCCGGCCCTAtaactccaggagTATATAAACTCAT
3	YGL098w/Use1:TMD_shuffled_2	ggttctggatcaggttcaAAGAGTAAATTGAGTATGTTAATCTTTATTCTGTTTGGATTCACTGTGATACTCATTGTTCAATTTTACATTATTTTGACATTCTTTTATCCGGCCCTAtaactccaggagTATATAAACTCAT
3	YHL031c/Gos1:length_22_V	ggttctggatcaggttcaAGAAGGAAGAAAAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACAGTTGTCGTAGTTTGGtaactccaggagTATATAAACTCAT
3	YHL031c/Gos1:length_24_V	ggttctggatcaggttcaAGAAGGAAGAAAAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACAGTTGTCGTAGTTGTCGTATGGtaactccaggagTATATAAACTCAT
3	YHL031c/Gos1:length_26_V	ggttctggatcaggttcaAGAAGGAAGAAAAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACAGTTGTCGTAGTTGTCGTAGTTGTCTGGtaactccaggagTATATAAACTCAT
3	YHL031c/Gos1:length_26_V_shuffled_1	ggttctggatcaggttcaAGAAGGAAGAAAAACTTTGTCACAGTAATAGTATTTTTGGTTGTCGTCATATTTACCCTGTTGGCGTTCACGTGTGCCGTTACCGTTGTACTTTGGtaactccaggagTATATAAACTCAT
3	YHL031c/Gos1:length_26_V_shuffled_2	ggttctggatcaggttcaAGAAGGAAGAAAAACTTTGCCGTTTTCACGCTGTTGTTTATATTTGTAGCGACCGTAGTCGTCATAGTTGTAACAACCCTTGTTTTGTGTCTTGTGGtaactccaggagTATATAAACTCAT

3	YHL031c/Gos1:length_26_VA	ggttctggatcaggttcaGAGAAGGAAGAAAAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACAGTTGCTGTCGCAGT AGCCGTTGCTTGGtaactccaggagTATATAAACTCAT
3	YIL004c/Bet1:length_24_V	ggttctggatcaggttcaGCCAGAAGATCTGGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTG TTGTCACAtaactccaggagTATATAAACTCAT
3	YIL004c/Bet1:length_26_V	ggttctggatcaggttcaGCCAGAAGATCTGGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTG TTGTCGTAGTTACAtaactccaggagTATATAAACTCAT
3	YIL004c/Bet1:length_26_V_shuffled_1	ggttctggatcaggttcaGCCAGAAGATCTGGGATTGTATTAGTATTTATAATCGTTAGTCTATTTGTTGGCGTATGGTTTGTGATAACATTTATGTGGTG GAAAATAGTCACAtaactccaggagTATATAAACTCAT
3	YIL004c/Bet1:length_26_V_shuffled_2	ggttctggatcaggttcaGCCAGAAGATCTGGGAGTATTTGGACAGTTATATTTTGGTTTTTTGTGGTCCTAGTAATGATAGGCTTTTTAGTAAAAATAGT AGTTTGGATCACAtaactccaggagTATATAAACTCAT
3	YIL004c/Bet1:length_26_VA	ggttctggatcaggttcaGCCAGAAGATCTGGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTG TTGCTGTCGCAACAtaactccaggagTATATAAACTCAT
3	YIL004c/Bet1:TMD_shuffled_1	ggttctggatcaggttcaGCCAGAAGATCTGGGACAGTAATTTGGTGGTTTAAAATGTTTGTGTTTATCCTAATATTTAGTATAATATTATGGGGCGTAAC Ataactccaggag TATATAAACTCAT
3	YIL004c/Bet1:TMD_shuffled_2	ggttctggatcaggttcaGCCAGAAGATCTGGGAGTATCTGGGTGATAATAAAATGGTTTATAATTTTTTGGTTTGGCGTATTAATGTTTCTAGTAACAAC Ataactccaggag TATATAAACTCAT
3	YIL065c/Fis1:length_22_V	ggttctggatcaggttcaAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAGTTGTC AGAtaactccaggagTATATAAACTCAT
3	YIL065c/Fis1:length_24_V	ggttctggatcaggttcaAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAGTTGTC GTAGTTAGAtaactccaggagTATATAAACTCAT
3	YIL065c/Fis1:length_26_V	ggttctggatcaggttcaAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAGTTGTC GTAGTTGTCGTAAGAtaactccaggagTATATAAACTCAT
3	YIL065c/Fis1:length_26_V_shuffled_1	ggttctggatcaggttcaAAGGAAACACTCAAGGCTGCCGCCGTCGCTGCTGTCGGTGTACTAGGCGTATTAGTTGTCAGTGTGGTAGGCGTTGTTGGA GTCTTCTTCGTGAGAtaactccaggagTATATAAACTCAT
3	YIL065c/Fis1:length_26_V_shuffled_2	ggttctggatcaggttcaAAGGAAACACTCAAGTTAAGTGTACTAGTTGCCGTGGTTGGCGGAGTCTTCGTCGTGGTAGGTGTAGCCGCTGTCGTTGTC TTCGCTGCTGGCAGAtaactccaggagTATATAAACTCAT
3	YIL065c/Fis1:length_26_VA	ggttctggatcaggttcaAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAGTTGCT GTCGCAGTAGCCAGAtaactccaggagTATATAAACTCAT
3	YIL065c/Fis1:TMD_shuffled_1	ggttctggatcaggttcaAAGGAAACACTCAAGGGACTAAGTGCTTTCGCCGCCTTCGTGGGCGTAGCTGTGGGTGTCGTCGGCGTTGCTTTAAGAAAC AAGAGAAGGtaactccaggagTATATAAACTCAT
3	YIL065c/Fis1:TMD_shuffled_2	ggttctggatcaggttcaAAGGAAACACTCAAGGTGTTCCTAGCTGCTGGCGCCTTAGCTGTCGCCGTCGGTGGCAGTGTTGTAGTGTTCGGAAGAAAC AAGAGAAGGtaactccaggagTATATAAACTCAT
3	YJL119c:length_22_V	ggttctggatcaggttcaACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTGTTGTCGTATTT taactccaggagTATATAAACTCAT
3	YJL119c:length_24_V	ggttctggatcaggttcaACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTGTTGTCGTAGTT GTCTTTtaactccaggagTATATAAACTCAT
3	YJL119c:length_26_V	ggttctggatcaggttcaACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTGTTGTCGTAGTT GTCGTAGTTTTTtaactccaggagTATATAAACTCAT
3	YJL119c:length_26_V_shuffled_1	ggttctggatcaggttcaACTGAAAAGCATGAAGTTTTCGTTAAGAAAATCGCCGTAGTCGTCGTGTTTGTAACTCTATATGTTTTTGCCGTCATAGTTTC TTTTAACGTATTTtaactccaggagTATATAAACTCAT
3	YJL119c:length_26_V_shuffled_2	ggttctggatcaggttcaACTGAAAAGCATGAAAAGTTTTATTTCGTAGCCGTTCTAGTGGTAAACATATTTACTGTCGTCTTTTCTGTTAAAGTTGTTGTA GCCATCGTCTTTtaactccaggagTATATAAACTCAT
3	YJL119c:length_26_VA	ggttctggatcaggttcaACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTGTTGCTGTCGC AGTAGCCGTTTTTtaactccaggagTATATAAACTCAT
3	YJL119c:TMD_shuffled_1	ggttctggatcaggttcaACTGAAAAGCATGAAGCCGTAACTATAATCAAATTTGTTTTTGCCTTCAAGGTGGTCTCTTTTTATCTAAACTTTTTTTTCGCAt aactccaggagTATATAAACTCAT
3	YJL119c:TMD_shuffled_2	ggttctggatcaggttcaACTGAAAAGCATGAAGCCCTATTTATAGCCGTTTCTACTGTATTTTATAAGAACGTCTTCATCAAATTTGTGTTTTTTTTCGCAt aactccaggagTATATAAACTCAT
3	YKL006c-A/Sft1:length_22_V	ggttctggatcaggttcaTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTGTTGTCAA Gtaactccaggag TATATAAACTCAT
3	YKL006c-A/Sft1:length_24_V	ggttctggatcaggttcaTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTGTTGTCGT AGTTAAGtaactccaggagTATATAAACTCAT
3	YKL006c-A/Sft1:length_26_V	ggttctggatcaggttcaTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTGTTGTCGT AGTTGTCGTAAAGtaactccaggagTATATAAACTCAT
3	YKL006c-A/Sft1:length_26_V_shuffled_1	ggttctggatcaggttcaTTAAAGGCTGGTAATCTGTGGGTCCTAATAGTCGGTTTGAGCTATGTTACCGTTTTCTTCATTGTGTTATTTTTAATGGTAATA AGAGCGGTAAAGtaactccaggagTATATAAACTCAT
3	YKL006c-A/Sft1:length_26_V_shuffled_2	ggttctggatcaggttcaTTAAAGGCTGGTAATATTGGTTTCGTATTATTTCTAATGATAACCGTCTTGCTGGTAGTTGTTTATGCGTGGTTCAGCTTAGTC AGAATAGTGAAGtaactccaggagTATATAAACTCAT
3	YKL006c-A/Sft1:length_26_VA	ggttctggatcaggttcaTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTGTTGCTGT CGCAGTAGCCAAGtaactccaggagTATATAAACTCAT
3	YKL006c-A/Sft1:TMD_shuffled_1	ggttctggatcaggttcaTTAAAGGCTGGTAATTGGTTACTGGCGTTCACCGTGTTAATATTCATTGGTAGCAGATTTCTAATATATTTGATGAAGTTATT Ttaactccaggag TATATAAACTCAT
3	YKL006c-A/Sft1:TMD_shuffled_2	ggttctggatcaggttcaTTAAAGGCTGGTAATTTCTATCTAATGACCATTTGGATATTTATAAGATTGTTCCTGGGTTTATTAAGCGTGGCGAAGTTATT Ttaactccaggag TATATAAACTCAT
3	YLR026c/Sed5:length_22_V	ggttctggatcaggttcaATAAAGAGTAATAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCGTTGTCAAT taactccaggag TATATAAACTCAT
3	YLR026c/Sed5:length_24_V	ggttctggatcaggttcaATAAAGAGTAATAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCGTTGTCGTA GTTAATtaactccaggagTATATAAACTCAT
3	YLR026c/Sed5:length_26_V	ggttctggatcaggttcaATAAAGAGTAATAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCGTTGTCGTA GTTGTCGTAAATtaactccaggagTATATAAACTCAT
3	YLR026c/Sed5:length_26_V_shuffled_1	ggttctggatcaggttcaATAAAGAGTAATAGAGTAGTTATATTCATTGTCGTTGTTTTATTTTTTTTTGTCGCCTTAGTCAAGTGGGTTTGGGCAGTAATC GTAGTTTTTAATtaactccaggagTATATAAACTCAT
3	YLR026c/Sed5:length_26_V_shuffled_2	ggttctggatcaggttcaATAAAGAGTAATAGAGTTGTCTTCGTTGTTGTAGTCTTTTTAATTATATTAGTTGTTATCTTTGTAGTAGTCTGGTTTTGGTTT GCAGCCAAGAATtaactccaggagTATATAAACTCAT
3	YLR026c/Sed5:length_26_VA	ggttctggatcaggttcaATAAAGAGTAATAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCGTTGCTGTC GCAGTAGCCAATtaactccaggagTATATAAACTCAT
3	YLR026c/Sed5:TMD_shuffled_1	ggttctggatcaggttcaATAAAGAGTAATAGAGTTTTTGTTATATGGATTTTATTAGTATTCGTCATCGCCGTTAAGTGGTTTTTTTTTTGCAAATtaactcca ggagTATATAAACTCAT
3	YLR026c/Sed5:TMD_shuffled_2	ggttctggatcaggttcaATAAAGAGTAATAGATTAGCAGTTTGGTTTATTATCTTTGTTGTTGTATGGTTTGTCTTATTTGCCAAGATATTCAATtaactcca ggagTATATAAACTCAT
3	YLR078c/Bos1:length_22_V	ggttctggatcaggttcaGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGGTTGTCGTAGTTAAA taactccaggag TATATAAACTCAT
3	YLR078c/Bos1:length_24_V	ggttctggatcaggttcaGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGGTTGTCGTAGTTGTC GTAAAAtaactccaggagTATATAAACTCAT
3	YLR078c/Bos1:length_26_V	ggttctggatcaggttcaGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGGTTGTCGTAGTTGTC GTAGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YLR078c/Bos1:length_26_V_shuffled_1	ggttctggatcaggttcaGTGTTCAAAGATAAACTCATTTTTGTTATCGTCTTAGTGGTTGTTATCTTGGTCGTAGTCTGGCTATATATATTGTATGCGGGT GTAATTGTCAAAtaactccaggagTATATAAACTCAT
3	YLR078c/Bos1:length_26_V_shuffled_2	ggttctggatcaggttcaGTGTTCAAAGATAAAGTTTATATAATTATTGCGGTCGTCGTTGTAGTCTTGCTATATTTGTGGCTCGTGTTAGTAGGTATCGT TATCGTCTTTAAAtaactccaggagTATATAAACTCAT
3	YLR078c/Bos1:length_26_VA	ggttctggatcaggttcaGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGGTTGCTGTCGCAGT AGCCGTTGCTAAAtaactccaggagTATATAAACTCAT

3	YLR078c/Bos1:TMD_shuffled_1	ggttctggatcaggttcaGTGTTCAAAGATAAAGGTGTGTATATTTGGTTGTTATTTATTATATTGTATCTACTCGTCGCGATCATCAAATGGTTAAGAtaactccaggagTATATAAACTCAT
3	YLR078c/Bos1:TMD_shuffled_2	ggttctggatcaggttcaGTGTTCAAAGATAAATATATACTATGGTTTATTATCATTATCTATCTCGTGTTGGTCGGTTTAGCGTTGAAATGGTTAAGAtaactccaggagTATATAAACTCAT
3	YLR093c/Nyv1:length_22_V	ggttctggatcaggttcaCAGAAGGTCAAAAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGGTTTGGtaactccaggagTATATAAACTCAT
3	YLR093c/Nyv1:length_24_V	ggttctggatcaggttcaCAGAAGGTCAAAAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGGTTGTCGTATGGtaactccaggagTATATAAACTCAT
3	YLR093c/Nyv1:length_26_V	ggttctggatcaggttcaCAGAAGGTCAAAAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGGTTGTCGTAGTTGTCTGGtaactccaggagTATATAAACTCAT
3	YLR093c/Nyv1:length_26_V_shuffled_1	ggttctggatcaggttcaCAGAAGGTCAAAAATGTATATACTATTTTTGTCACTGCTGCTGTTCTGTTATTACTATTCATGGTTATAAGTGTCATTGTAACGTTCTTTTTCTGGtaactccaggagTATATAAACTCAT
3	YLR093c/Nyv1:length_26_V_shuffled_2	ggttctggatcaggttcaCAGAAGGTCAAAAATGTCACGATGTTAACTTATACTGCTTTATTCTTTCTAATTGTTATAGTTGTAGTCGTATTCAGTGCTTTTATTTTCCTGTGGtaactccaggagTATATAAACTCAT
3	YLR093c/Nyv1:length_26_VA	ggttctggatcaggttcaCAGAAGGTCAAAAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGGTTGCTGTCGCAGTATGGtaactccaggagTATATAAACTCAT
3	YLR093c/Nyv1:TMD_shuffled_1	ggttctggatcaggttcaCAGAAGGTCAAAAATCTGTATAGTGCTTTATTCGCTATTTTCTTCATATTTACGTTAATTACTCTATTTACTATGGTATGGtaactccaggagTATATAAACTCAT
3	YLR093c/Nyv1:TMD_shuffled_2	ggttctggatcaggttcaCAGAAGGTCAAAAATATTATGACTACGTTCTATTTTTTCGCTGTATTCTTAATTCTACTGAGTACTATAGCTTTTTTATGGtaactccaggagTATATAAACTCAT
3	YLR238w/Far10:length_22_V	ggttctggatcaggttcaACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTGTTTCCtaactccaggagTATATAAACTCAT
3	YLR238w/Far10:length_24_V	ggttctggatcaggttcaACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTGTTGTCGTATCCtaactccaggagTATATAAACTCAT
3	YLR238w/Far10:length_26_V	ggttctggatcaggttcaACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTGTTGTCGTAGTTGTCTCCtaactccaggagTATATAAACTCAT
3	YLR238w/Far10:length_26_V_shuffled_1	ggttctggatcaggttcaACGCTATGTAATCATGTAATTAAGGCTCTTGTTTTCGGAATCATTTTCGGGACTTTAATTACAGTTTTTATATCCACAGTCGTCGTCATCCTATCCtaactccaggagTATATAAACTCAT
3	YLR238w/Far10:length_26_V_shuffled_2	ggttctggatcaggttcaACGCTATGTAATCATTTAACAGTAGCTATTCTAGTTTTCATCGTCGGGGTCAAGATTTTCATAGGAACATTTGTCGTTTCCACTCTTATTATCTCCtaactccaggagTATATAAACTCAT
3	YLR238w/Far10:length_26_VA	ggttctggatcaggttcaACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTGTTGCTGTCGCAGTATCCtaactccaggagTATATAAACTCAT
3	YLR238w/Far10:TMD_shuffled_1	ggttctggatcaggttcaACGCTATGTAATCATATAATCATCACTGTCCTTATTTCCGGACTATTCAAGTTTACAGGGTTATTCATTGCTATTACATCCCCCAACtaactccaggagTATATAAACTCAT
3	YLR238w/Far10:TMD_shuffled_2	ggttctggatcaggttcaACGCTATGTAATCATTTTTTCATCGTCAAGCTTATATCCGGATTAACTTTCGCTCTAATTACAACAATCGGGATTATTTCCCCCAACtaactccaggagTATATAAACTCAT
3	YLR268w/Sec22:TMD_shuffled_1	ggttctggatcaggttcaGCGCAAAAGATCAACATTTTCTATTTCGCTATCCCTATTTTTTTCGTCCAAATCAGTGTCGATGTCGCTCTCTTCTTCCTCTTTTTGTGGTGGCTCAAAtaactccaggagTATATAAACTCAT
3	YLR268w/Sec22:TMD_shuffled_2	ggttctggatcaggttcaGCGCAAAAGATCAACCTCGCTTGGCAAATTATCTGGGCTCCTTATATTATCGTCCTCAGTGTCTTGTTTTTCCTCTTCGATTTCTTCTTTGTCTTCAAAtaactccaggagTATATAAACTCAT
3	YMR161w/Hlj1:length_22_V	ggttctggatcaggttcaTCGCAATTAAAAAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTGTTGTCGTAAGTtaactccaggagTATATAAACTCAT
3	YMR161w/Hlj1:length_24_V	ggttctggatcaggttcaTCGCAATTAAAAAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTGTTGTCGTAGTTGTCAGTtaactccaggagTATATAAACTCAT
3	YMR161w/Hlj1:length_26_V	ggttctggatcaggttcaTCGCAATTAAAAAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTGTTGTCGTAGTTGTCGTAGTTAGTtaactccaggagTATATAAACTCAT
3	YMR161w/Hlj1:length_26_V_shuffled_1	ggttctggatcaggttcaTCGCAATTAAAAAATGTACCTATTGTAGTTAAAGTCATGCTGCTTGTCGTTATTCTCGTTTACATCATCCTTGATTTCATGTTTGTTTTTGTTAGTtaactccaggagTATATAAACTCAT
3	YMR161w/Hlj1:length_26_V_shuffled_2	ggttctggatcaggttcaTCGCAATTAAAAAATGTTCCTGTTCTGGTACTCTTTGTCCTTGTTATTATGAAACTTGATGTCTACTTTATCTTCGTAATGGTTATCGTTATTAGTtaactccaggagTATATAAACTCAT
3	YMR161w/Hlj1:length_26_VA	ggttctggatcaggttcaTCGCAATTAAAAAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTGTTGCTGTCGCAGTAGCCGTTAGTtaactccaggagTATATAAACTCAT
3	YMR161w/Hlj1:TMD_shuffled_1	ggttctggatcaggttcaTCGCAATTAAAAAATATCGTTTTTTTCATTGTTATGCTGATGCTCCCTCTTGATTACAAAATCTTTCTTATTAGTtaactccaggagTATATAAACTCAT
3	YMR161w/Hlj1:TMD_shuffled_2	ggttctggatcaggttcaTCGCAATTAAAAAATCTTCTTATCCTCCTGTTTGTTCCTGATATGATTAAATTTGTTTTCATCATTATGTACAGTtaactccaggagTATATAAACTCAT
3	YMR183c/Sso2:length_26_V	ggttctggatcaggttcaGCAAGAAAAAACAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTGGTTGTCGAAtaactccaggagTATATAAACTCAT
3	YMR183c/Sso2:length_26_V_shuffled_1	ggttctggatcaggttcaGCAAGAAAAAACAAAGTTGTTGCTTGTATTATCTTTGTGATCATCTCCATATTTCCAGTTGTTATTGTCGTCGTTGTGTTGGTTGTTTGCAGAGAAtaactccaggagTATATAAACTCAT
3	YMR183c/Sso2:length_26_V_shuffled_2	ggttctggatcaggttcaGCAAGAAAAAACAAAGTGATCGTTGTCCCAATTGTTGTTTTTGTTTTGATCTTTGTGATAGTTATCAGAGTTATTTCCTGTGTTTGCGTCGCTGAAtaactccaggagTATATAAACTCAT
3	YMR183c/Sso2:length_26_VA	ggttctggatcaggttcaGCAAGAAAAAACAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTGGTTGCTGAAtaactccaggagTATATAAACTCAT
3	YMR183c/Sso2:TMD_shuffled_1	ggttctggatcaggttcaGCAAGAAAAAACAAACCATGCGTTGTTATATTGGTTGCTGTTGTTTTTTTTTCCGTGATCGTCGTGATCGTTATTTGTATTAGAATCGAAACAAGAAAGtaactccaggagTATATAAACTCAT
3	YMR183c/Sso2:TMD_shuffled_2	ggttctggatcaggttcaGCAAGAAAAAACAAATCCCCAATCATTTGTGTTATCGTGAGAATCGTTATTGCTTTTGTTATAGTTTGCTTGGTTGTTTTTGTGGTCGAAACAAGAAAGtaactccaggagTATATAAACTCAT
3	YMR197c/Vti1:length_22_V	ggttctggatcaggttcaCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAGTTGTCAAGtaactccaggagTATATAAACTCAT
3	YMR197c/Vti1:length_24_V	ggttctggatcaggttcaCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAGTTGTCGTAGTTAAGtaactccaggagTATATAAACTCAT
3	YMR197c/Vti1:length_26_V	ggttctggatcaggttcaCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAGTTGTCGTAGTTGTCGTAAAGtaactccaggagTATATAAACTCAT
3	YMR197c/Vti1:length_26_V_shuffled_1	ggttctggatcaggttcaCTAGTTGCTAATAAATTGGTTGTAGTCGCAATAATTATCTTCATAGTCTTGGTTTTGATTCTAGTAGCCGTCTTATCATATCTTTTCGTTAGCAAGtaactccaggagTATATAAACTCAT
3	YMR197c/Vti1:length_26_V_shuffled_2	ggttctggatcaggttcaCTAGTTGCTAATAAAAGCGTTTTCTTATTGCTTGTCGTTATAGTTGTCTTCTCAGTACTAGCATTGATTTATGTCGTAATTATCGCCTTGATAAAGtaactccaggagTATATAAACTCAT
3	YMR197c/Vti1:length_26_VA	ggttctggatcaggttcaCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAGTTGCTGTCGCAGTAGCCAAGtaactccaggagTATATAAACTCAT
3	YMR197c/Vti1:TMD_shuffled_1	ggttctggatcaggttcaCTAGTTGCTAATAAAGTTAGCATACTTGCCTTCTTCATAATTTTACTAATCGTCATTGCATATTTGTCATTGTTGAAGTTTAAAtaactccaggagTATATAAACTCAT
3	YMR197c/Vti1:TMD_shuffled_2	ggttctggatcaggttcaCTAGTTGCTAATAAATTCCTTTTAATTATTTCATTGTTGGCCATACTAGTCGTTTATGCAATATTCATCTTGAGCAAGTTTAAAtaactccaggagTATATAAACTCAT
3	YNL070w/Tom7:length_24_V	ggttctggatcaggttcaGAACGTATTTCCAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGTTGTCGCAtaactccaggagTATATAAACTCAT
3	YNL070w/Tom7:length_26_V	ggttctggatcaggttcaGAACGTATTTCCAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGTTGTCGTAGTTGCAtaactccaggagTATATAAACTCAT
3	YNL070w/Tom7:length_26_V_shuffled_1	ggttctggatcaggttcaGAACGTATTTCCAAATGGGTTCATTATGTATTGACTTATGGCATCGGCACTATTTTGTTATTTGTTGTATGGTTGGCACATAATCCAGTCGTTGCAtaactccaggagTATATAAACTCAT
3	YNL070w/Tom7:length_26_V_shuffled_2	ggttctggatcaggttcaGAACGTATTTCCAAAGTTGTAGCAGTATTGATCGTTTTATATGTTACTTGGCATATTTTGTTGCCAGGCTGGGTCCATTTTACTTATGGCAATGCAtaactccaggagTATATAAACTCAT

3	YNL070w/Tom7:length_26_VA	ggttctggatcaggttcaGAACGTATTTCCAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGTTGCTGTCGCAGCAtaactccaggagTATATAAACTCAT
3	YNL070w/Tom7:TMD_shuffled_1	ggttctggatcaggttcaGAACGTATTTCCAAAGTACATTGGATCCATTTGTATTATTTAACTTTGTTTGTTGCATTGTGGGGCGGCACTAATCCAATTGCACACACTTCTAATtaactccaggagTATATAAACTCAT
3	YNL070w/Tom7:TMD_shuffled_2	ggttctggatcaggttcaGAACGTATTTCCAAATTTCATATTAATCCATGGTGGTTAGGCATCTATACTGTACATTTGGTTTTGGGCTATTTGACTGCAGCACACACTTCTAATtaactccaggagTATATAAACTCAT
3	YNL111c/Cyb5:length_22_V	ggttctggatcaggttcaCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACGTTGTCGAAtaactccaggagTATATAAACTCAT
3	YNL111c/Cyb5:length_24_V	ggttctggatcaggttcaCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACGTTGTCGTAGTTGAAtaactccaggagTATATAAACTCAT
3	YNL111c/Cyb5:length_26_V	ggttctggatcaggttcaCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACGTTGTCGTAGTTGTCGTAGAAtaactccaggagTATATAAACTCAT
3	YNL111c/Cyb5:length_26_V_shuffled_1	ggttctggatcaggttcaCAAAGTAAAGGTAGTAACTATGTTGGTGTTGTTGCCACAGTCGTTTTGGTCGTAATTATGTTAGTCGCTTATGGTGTACTATTGTTGTTGATAGAAtaactccaggagTATATAAACTCAT
3	YNL111c/Cyb5:length_26_V_shuffled_2	ggttctggatcaggttcaCAAAGTAAAGGTAGTCTAGTTTTGGGTGCCATTAACGTCGCTGTTGTTGTATTGGTCTATACAATATTGTTAGTCATGGTTGGTTTATGTATTGGAAtaactccaggagTATATAAACTCAT
3	YNL111c/Cyb5:length_26_VA	ggttctggatcaggttcaCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACGTTGCTGTCGCAGTAGCCGAAtaactccaggagTATATAAACTCAT
3	YNL111c/Cyb5:TMD_shuffled_1	ggttctggatcaggttcaCAAAGTAAAGGTAGTTTGTATATTGTTGCTATGTATGTTTTGAACATATTATTGGTCCTAACAGCCGGTTTGGGTGAAtaactccaggagTATATAAACTCAT
3	YNL111c/Cyb5:TMD_shuffled_2	ggttctggatcaggttcaCAAAGTAAAGGTAGTAACGTCTTGGCTATGTTGTTGATAGTTCTAGGTTTGGCCTATACAGTTTTATATATTGGTGAAtaactccaggagTATATAAACTCAT
3	YNL131w/Tom22:length_24_V	ggttctggatcaggttcaACAAAATCCGGAAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGTTGTCGAAtaactccaggagTATATAAACTCAT
3	YNL131w/Tom22:length_26_V	ggttctggatcaggttcaACAAAATCCGGAAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGTTGTCGTAGTTGAAtaactccaggagTATATAAACTCAT
3	YNL131w/Tom22:length_26_V_shuffled_1	ggttctggatcaggttcaACAAAATCCGGAAACTTGCTTATACCACTCTTATTGGTCTCCACCGTTTCTACCCTAGTACTTGTGACTGCTGCCTGGGTTTTAGCTACTGGTGAAtaactccaggagTATATAAACTCAT
3	YNL131w/Tom22:length_26_V_shuffled_2	ggttctggatcaggttcaACAAAATCCGGAAACATACTATCTCTTGCTGGTGTCTTAACTTTATTGTGGACCGCCGTGGCTTTGGTTACTCCACTCGTTTCCCTTACCGTAGAAtaactccaggagTATATAAACTCAT
3	YNL131w/Tom22:length_26_VA	ggttctggatcaggttcaACAAAATCCGGAAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGTTGCTGTCGCAGAAtaactccaggagTATATAAACTCAT
3	YNL131w/Tom22:TMD_shuffled_1	ggttctggatcaggttcaACAAAATCCGGAAACTGGTCCGGTACTACCGCCCCATCTACCCTTTTATTGTTGCTTGTGGCTATAACTCTACTCGCTTTAGAACAACAGCTAATCtaactccaggagTATATAAACTCAT
3	YNL131w/Tom22:TMD_shuffled_2	ggttctggatcaggttcaACAAAATCCGGAAACCTAGCTTTATTGACCTTGCTTTCCCCAGCTGCCACCTCTATAACTTTACTCGGTTGGCTTACTGTGGAACAACAGCTAATCtaactccaggagTATATAAACTCAT
3	YNL188w/Kar1:length_22_V	ggttctggatcaggttcaAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATGTTGTCAGGtaactccaggagTATATAAACTCAT
3	YNL188w/Kar1:length_24_V	ggttctggatcaggttcaAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATGTTGTCGTAGTTAGGtaactccaggagTATATAAACTCAT
3	YNL188w/Kar1:length_26_V	ggttctggatcaggttcaAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATGTTGTCGTAGTTGTCGTAAGGtaactccaggagTATATAAACTCAT
3	YNL188w/Kar1:length_26_V_shuffled_1	ggttctggatcaggttcaAAAAAGTATAGGGAATTATATTGCACAGTCTATATTGTGTGGATTTTATTGATATATGTCGTATTCGTTTATTTATATTGTGTAATAGTTAATAGGtaactccaggagTATATAAACTCAT
3	YNL188w/Kar1:length_26_V_shuffled_2	ggttctggatcaggttcaAAAAAGTATAGGGAATTATTGTATGTCACAGTCTATATTGTAATTGTTTTATATTATATATGTGTGTGGGTTTTAGTATATTTCATAAATTGCAGGtaactccaggagTATATAAACTCAT
3	YNL188w/Kar1:length_26_VA	ggttctggatcaggttcaAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATGTTGCTGTCGCAGTAGCCAGGtaactccaggagTATATAAACTCAT
3	YNL188w/Kar1:TMD_shuffled_1	ggttctggatcaggttcaAAAAAGTATAGGGAATGCTATGTGTGTTATTATATTTTGTGGATATATTTAATTACATTAATATTAAATTTCTATAGGTTTtaactccaggagTATATAAACTCAT
3	YNL188w/Kar1:TMD_shuffled_2	ggttctggatcaggttcaAAAAAGTATAGGGAAACATGGTATTATTTGATAATTGTGATAAATTTATATTTATGCTTATATTTCTATATTTGTAGGTTTtaactccaggagTATATAAACTCAT
3	YOL018c/Tlg2:length_22_V	ggttctggatcaggttcaACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGGTTGTCGTAGTTAAAtaactccaggagTATATAAACTCAT
3	YOL018c/Tlg2:length_24_V	ggttctggatcaggttcaACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGGTTGTCGTAGTTGTCGTAAAAtaactccaggagTATATAAACTCAT
3	YOL018c/Tlg2:length_26_V	ggttctggatcaggttcaACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGGTTGTCGTAGTTGTCGTAGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YOL018c/Tlg2:length_26_V_shuffled_1	ggttctggatcaggttcaACTCAAAAATGTAAAGTCTTGTGCGTAGCGGTCTTATTCGTCCTCTTGGTTCTAGTAATAGTTTTTATGGTAGTCTTAGTTACATTTGTTATTAAAtaactccaggagTATATAAACTCAT
3	YOL018c/Tlg2:length_26_V_shuffled_2	ggttctggatcaggttcaACTCAAAAATGTAAAATATTATTACTCGTATTGGTTTGCTTTCTAGTTGTTGTCATGGTATTCTTTGCGGTAGTCGTCTTGACAATTGTCGTTAAAtaactccaggagTATATAAACTCAT
3	YOL018c/Tlg2:length_26_VA	ggttctggatcaggttcaACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGGTTGCTGTCGCAGTAGCCGTTGCTAAAtaactccaggagTATATAAACTCAT
3	YOL018c/Tlg2:TMD_shuffled_1	ggttctggatcaggttcaACTCAAAAATGTAAATTGGTATTATTTTTAGTCCTCACATTCTGCTTTGTTATATTGCTAGCGATTATGAAACCACATGGCGGTtaactccaggagTATATAAACTCAT
3	YOL018c/Tlg2:TMD_shuffled_2	ggttctggatcaggttcaACTCAAAAATGTAAAGTAATATTATTACTAATGGTCTTCGCGTGCTTTATTCTCTTGTTGTTTGTTACAAAACCACATGGCGGTtaactccaggagTATATAAACTCAT
3	YOL044w/Pex15:length_22_V	ggttctggatcaggttcaCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGGTTGTCGTAGTTAAAtaactccaggagTATATAAACTCAT
3	YOL044w/Pex15:length_24_V	ggttctggatcaggttcaCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGGTTGTCGTAGTTGTCGTAAAAtaactccaggagTATATAAACTCAT
3	YOL044w/Pex15:length_26_V	ggttctggatcaggttcaCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGGTTGTCGTAGTTGTCGTAGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YOL044w/Pex15:length_26_V_shuffled_1	ggttctggatcaggttcaCATTTCACAAGAAGTGTTTGTGTTGTCGTAGTTCTCGTCCTGCTCAACCTGGTATTGGTCCTGGGACTTACACTAGGTTTAGTGAACCTGAAAAAAtaactccaggagTATATAAACTCAT
3	YOL044w/Pex15:length_26_V_shuffled_2	ggttctggatcaggttcaCATTTCACAAGAAGTCTTAAACTCCTGCTCAACACACTGCTAGTGAACGTCTGTGTATTATTGGGACTGGTTGTCGTAGTCGGTGTTGTTCTGAAAtaactccaggagTATATAAACTCAT
3	YOL044w/Pex15:length_26_VA	ggttctggatcaggttcaCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGGTTGCTGTCGCAGTAGCCGTTGCTAAAtaactccaggagTATATAAACTCAT
3	YOL044w/Pex15:TMD_shuffled_1	ggttctggatcaggttcaCATTTCACAAGAAGTCTCACACTGGGAAAACTGAACCTGTTGTGTCTACTTTTAGTGCTGCTCGGTAACAAAAAATATAAGTCAtaactccaggagTATATAAACTCAT
3	YOL044w/Pex15:TMD_shuffled_2	ggttctggatcaggttcaCATTTCACAAGAAGTCTAGTGGGACTGCTGTTATTGACACTGAACCTCCTTCTCTGTGGTAAACTGAACAAAAAATATAAGTCAtaactccaggagTATATAAACTCAT
3	YOR036w/Pep12:length_22_V	ggttctggatcaggttcaAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YOR036w/Pep12:length_24_V	ggttctggatcaggttcaAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGGTTGTCGTAGTTAAAtaactccaggagTATATAAACTCAT
3	YOR036w/Pep12:length_26_V	ggttctggatcaggttcaAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGGTTGTCGTAGTTGTCGTAAAAtaactccaggagTATATAAACTCAT
3	YOR036w/Pep12:length_26_V_shuffled_1	ggttctggatcaggttcaAAACGTACGAGCAGACTCATTTTTAGGGTATATCTTGTTGTCCTTTTTTGGGTAATGGTCATTATGTTGATTTTGCTCGTGCTTGTAGTGGTTAAAtaactccaggagTATATAAACTCAT

3	YOR036w/Pep12:length_26_V_shuffled_2	ggttctggatcaggttcaAAACGTACGAGCAGACTCTATTGGGTAGTGAGGGTCGTGGTAATTCTTATTGTTCTTTTGGTCTTTCTCTTTCTTGTAATGATTATGTTGGTTAAAtaactccaggagTATATAAACTCAT
3	YOR036w/Pep12:length_26_VA	ggttctggatcaggttcaAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGGTTGCTGTCGCAGTAGCCAAAtaactccaggagTATATAAACTCAT
3	YOR036w/Pep12:TMD_shuffled_1	ggttctggatcaggttcaAAACGTACGAGCAGAATGGTGTTTCTTTTGGTGTATCTCTTGCTTATGGTAATTCTTATTTTTTGGATTCTCAGGAAATTGtaactccaggagTATATAAACTCAT
3	YOR036w/Pep12:TMD_shuffled_2	ggttctggatcaggttcaAAACGTACGAGCAGATTTTTGGTACTCTTTATTCTCCTTCTTATGCTTAGGATTTGGGTGTATGTGTTGATGATTAAATTGtaactccaggagTATATAAACTCAT
3	YOR045w/Tom6:length_22_V	ggttctggatcaggttcaGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTGTTGTCGTACAAtaactccaggagTATATAAACTCAT
3	YOR045w/Tom6:length_24_V	ggttctggatcaggttcaGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTGTTGTCGTAGTTGTCCAAtaactccaggagTATATAAACTCAT
3	YOR045w/Tom6:length_26_V	ggttctggatcaggttcaGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTGTTGTCGTAGTTGTCGTAGTTCAATaactccaggagTATATAAACTCAT
3	YOR045w/Tom6:length_26_V_shuffled_1	ggttctggatcaggttcaGCTTTCAAGGAATCTGTCTACCTAGCGTTCGCCGTATTTGCAGTTAACGTTTTCCCAGGTGTAATTGGTATTGTCGTTGTTACACTAGTTGCACAATaactccaggagTATATAAACTCAT
3	YOR045w/Tom6:length_26_V_shuffled_2	ggttctggatcaggttcaGCTTTCAAGGAATCTGTATTCGTTGCAGGTATTGTAGGTGCCTACGTCATTACAGCATTTAACGCGGTCCTATTCGTTGTTCCAGTTCTAGTTCAATaactccaggagTATATAAACTCAT
3	YOR045w/Tom6:length_26_VA	ggttctggatcaggttcaGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTGTTGCTGTCGCAGTAGCCGTTCAATaactccaggagTATATAAACTCAT
3	YOR045w/Tom6:TMD_shuffled_1	ggttctggatcaggttcaGCTTTCAAGGAATCTACAGCACCATTCATTGCCTTCGTTAACGGTATTGCACTACTATACTTTGTTGCGGGTCAATCTCCACTCATGtaactccaggagTATATAAACTCAT
3	YOR045w/Tom6:TMD_shuffled_2	ggttctggatcaggttcaGCTTTCAAGGAATCTCTAAACGCAACAATTGGTGGTCTATTCGTTTTTCCATTCGCGGTTGCATACGCCATTCAATCTCCACTCATGtaactccaggagTATATAAACTCAT
3	YOR075w/Ufe1:length_22_V	ggttctggatcaggttcaGGAAGAACTGCTAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGTTGTCGTAGGTtaactccaggagTATATAAACTCAT
3	YOR075w/Ufe1:length_24_V	ggttctggatcaggttcaGGAAGAACTGCTAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGTTGTCGTAGTTGTCGGTtaactccaggagTATATAAACTCAT
3	YOR075w/Ufe1:length_26_V	ggttctggatcaggttcaGGAAGAACTGCTAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGTTGTCGTAGTTGTCGTAGTTGGTtaactccaggagTATATAAACTCAT
3	YOR075w/Ufe1:length_26_V_shuffled_1	ggttctggatcaggttcaGGAAGAACTGCTAAACTATTCATGATTATAGTTTATGTTGTAACCGATGTCACTTTGATCATGTATTTTGGTGTAGGTGTTGCCGTCGTAGTTGGTtaactccaggagTATATAAACTCAT
3	YOR075w/Ufe1:length_26_V_shuffled_2	ggttctggatcaggttcaGGAAGAACTGCTAAATTCTATATTGTTATGGTCACTTTTGGTGTTATCGTTGTCGTTTATTTGACCCTAATGGATGGTATAGTAGTAGCCGTAGGTtaactccaggagTATATAAACTCAT
3	YOR075w/Ufe1:length_26_VA	ggttctggatcaggttcaGGAAGAACTGCTAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGTTGCTGTCGCAGTAGCCGTTGGTtaactccaggagTATATAAACTCAT
3	YOR075w/Ufe1:TMD_shuffled_1	ggttctggatcaggttcaGGAAGAACTGCTAAAACTTTTACCGTTTTCGGTATTTATATATATATGGCCATCGGTGTAATGCTATTGGATGGTtaactccaggagTATATAAACTCAT
3	YOR075w/Ufe1:TMD_shuffled_2	ggttctggatcaggttcaGGAAGAACTGCTAAATTTGTTGGTATGTATTATGGTTTGCTAACTGTAATGGATACCTTCATTGCCATAATCGGTtaactccaggagTATATAAACTCAT
3	YOR106w/Vam3:length_22_V	ggttctggatcaggttcaAACAAATGCGGTAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAGTTGTCGTAGTTAGTtaactccaggagTATATAAACTCAT
3	YOR106w/Vam3:length_24_V	ggttctggatcaggttcaAACAAATGCGGTAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAGTTGTCGTAGTTGTCGTAAGTtaactccaggagTATATAAACTCAT
3	YOR106w/Vam3:length_26_V	ggttctggatcaggttcaAACAAATGCGGTAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAGTTGTCGTAGTTGTCGTAGTTGTCAGTtaactccaggagTATATAAACTCAT
3	YOR106w/Vam3:length_26_V_shuffled_1	ggttctggatcaggttcaAACAAATGCGGTAAGGTAGTGGTATTGCTTGTCGTCATAGTTGTTATAGTCGTCTGCCTATTAGTAACCATTGTTGTGATGGTAGCTGTTATCAGTtaactccaggagTATATAAACTCAT
3	YOR106w/Vam3:length_26_V_shuffled_2	ggttctggatcaggttcaAACAAATGCGGTAAGGTAGTACTTGCTGTGTTAATAGTGTGCTTGATAATGGTCGTTATCACCGTTCTAATTGTCGTCGTCGTTGTAGTAGTTAGTtaactccaggagTATATAAACTCAT
3	YOR106w/Vam3:length_26_VA	ggttctggatcaggttcaAACAAATGCGGTAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAGTTGCTGTCGCAGTAGCCGTTGCTAGTtaactccaggagTATATAAACTCAT
3	YOR324c/Frt1:length_24_V	ggttctggatcaggttcaACGAAATCCATCTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCGTTGTCAAGtaactccaggagTATATAAACTCAT
3	YOR324c/Frt1:length_26_V	ggttctggatcaggttcaACGAAATCCATCTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCGTTGTCGTAGTTAAGtaactccaggagTATATAAACTCAT
3	YOR324c/Frt1:length_26_V_shuffled_1	ggttctggatcaggttcaACGAAATCCATCTACCTTGCATTTTATTTGATCGGCGTCATTGTCGTTTATGTTCTGAGGATGATTTTTATTATCGACGGCGTTTACTTCGTAAAGtaactccaggagTATATAAACTCAT
3	YOR324c/Frt1:length_26_V_shuffled_2	ggttctggatcaggttcaACGAAATCCATCTACTACATGGTCATTTTTTATTTGAGGGTTGCAGGCCTGTATGTAATCGTCGTTCTTATCGACTTTGGCATTTTCATTGTTAAGtaactccaggagTATATAAACTCAT
3	YOR324c/Frt1:length_26_VA	ggttctggatcaggttcaACGAAATCCATCTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCGTTGCTGTCGCAAAGtaactccaggagTATATAAACTCAT
3	YOR324c/Frt1:TMD_shuffled_1	ggttctggatcaggttcaACGAAATCCATCTACGCAATCATTCTTCTGATTATGAGGGGCTTTGGCTACGTCGTTTTTATCTATTTCATTTTGGACTATAAGAATTTGTTAACTtaactccaggagTATATAAACTCAT
3	YOR324c/Frt1:TMD_shuffled_2	ggttctggatcaggttcaACGAAATCCATCTACATCCTGTACAGGTTCGACCTTTTTATCATTGCAATGGTTGGCGGCTATTTTTATATTTTGATTGTCAAGAATTTGTTAACTtaactccaggagTATATAAACTCAT
3	YOR327c/Snc2:length_24_V	ggttctggatcaggttcaGATCTAAAAATGAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCGTTGTCAGCtaactccaggagTATATAAACTCAT
3	YOR327c/Snc2:length_26_V	ggttctggatcaggttcaGATCTAAAAATGAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCGTTGTCGTAGTTAGCtaactccaggagTATATAAACTCAT
3	YOR327c/Snc2:length_26_V_shuffled_1	ggttctggatcaggttcaGATCTAAAAATGAGATTCATCGTCGTTGTTTGTATTATGGTCGTCTTAGTTATCGTACATCCTATTCTAGTTTTAGTGTTCGTAATTTTAGTTAGCtaactccaggagTATATAAACTCAT
3	YOR327c/Snc2:length_26_V_shuffled_2	ggttctggatcaggttcaGATCTAAAAATGAGACATTTAATTTTAATCGTTGTCGTTTGTGTCATCGTGGTACTACCTATTGTTTTCATTGTCTTCTTAGTTGTTTGTAATGAGCtaactccaggagTATATAAACTCAT
3	YOR327c/Snc2:length_26_VA	ggttctggatcaggttcaGATCTAAAAATGAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCGTTGCTGTCGCAAGCtaactccaggagTATATAAACTCAT
3	YOR327c/Snc2:TMD_shuffled_1	ggttctggatcaggttcaGATCTAAAAATGAGAGTGTTAGTACATGTCTTCATTTTCGTCGTTATTGTTTGTATCCTATTAGTTATTCCTTTAATGATCAGCtaactccaggagTATATAAACTCAT
3	YOR327c/Snc2:TMD_shuffled_2	ggttctggatcaggttcaGATCTAAAAATGAGATTAGTCTTATGTCTAGTAATTTTAGTTGTCATGATCGTTATCGTGCCTCATGTTTTCATTTTCATTAGCtaactccaggagTATATAAACTCAT
3	YPL192c/Prm3:length_26_V	ggttctggatcaggttcaGAGAATAAGGGTTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YPL192c/Prm3:length_26_V_shuffled_1	ggttctggatcaggttcaGAGAATAAGGGTTCACTAGTAGGTAACGTCTTAGGTCTTGCTTTCTACGTTTCGGTTGCCTTTACTGTTTTCAGTGCTATTACTCAAGGTGCAAAAtaactccaggagTATATAAACTCAT
3	YPL192c/Prm3:length_26_V_shuffled_2	ggttctggatcaggttcaGAGAATAAGGGTTCAGCTTCGTTCTTATTCGTTTTTACTGCTGGTACTCTAGTTGTAGGTGCAAACGCCTACGTTCAACTTAGTGGTGTCATTAAAtaactccaggagTATATAAACTCAT
3	YPL192c/Prm3:length_26_VA	ggttctggatcaggttcaGAGAATAAGGGTTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTGTTGCTAAAtaactccaggagTATATAAACTCAT
3	YPL192c/Prm3:TMD_shuffled_1	ggttctggatcaggttcaGAGAATAAGGGTTCAGGTGCTCAAGTTGGTGGTATTGCTAACTTAAGTTTCGCCCTTTACTCGGTAGTTACTACTTTTCTATTCGCAAAAGCTCTTCAAAATtaactccaggagTATATAAACTCAT
3	YPL192c/Prm3:TMD_shuffled_2	ggttctggatcaggttcaGAGAATAAGGGTTCACTAGTTTTAAACGCATTTCAAGGTGGTTACATTGCTGCTTTCGGTGCCTCGACTTTCAGTGTAACTCTTGTTAAAGCTCTTCAAAATtaactccaggagTATATAAACTCAT

3	YPL200w/Csm4:length_22_V	ggttctggatcaggttcaTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGTTGTCGTAGAAtaactccaggagTATATAAACTCAT
3	YPL200w/Csm4:length_24_V	ggttctggatcaggttcaTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGTTGTCGTAGTTGTCGAAtaactccaggagTATATAAACTCAT
3	YPL200w/Csm4:length_26_V	ggttctggatcaggttcaTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGTTGTCGTAGTTGTCGTAGTTGAAtaactccaggagTATATAAACTCAT
3	YPL200w/Csm4:length_26_V_shuffled_1	ggttctggatcaggttcaTTCGTCATCGCGGAAGTCTCAGTTTTTTTAATAGTAGTATTGATTATATTGATATCTTTTTCGGTTGTTGTCTTTCTAAATGTTGTATGGGTAGAAtaactccaggagTATATAAACTCAT
3	YPL200w/Csm4:length_26_V_shuffled_2	ggttctggatcaggttcaTTCGTCATCGCGGAATTTTCTTTAATTCTAGTCGTAGTAGTTATAGTAGTTTTGAATTTGGTTTTTGTCGTTATATCAGTATGGTCGTTTATAGAAtaactccaggagTATATAAACTCAT
3	YPL200w/Csm4:length_26_VA	ggttctggatcaggttcaTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGTTGCTGTCGCAGTAGCCGTTGAAtaactccaggagTATATAAACTCAT
3	YPL200w/Csm4:TMD_shuffled_1	ggttctggatcaggttcaTTCGTCATCGCGGAAATATTGTTTTTAGTATCTATTGTAATATTGTTTTGGTCATTTGTTCTAAATTCGATAGAAGTTtaactccaggagTATATAAACTCAT
3	YPL200w/Csm4:TMD_shuffled_2	ggttctggatcaggttcaTTCGTCATCGCGGAATTTGTTTTAATAATATCAGTATTTTGGTCGTCTATAAATCTAGTATTGTTTATTTTGGAAGTTtaactccaggagTATATAAACTCAT
3	YPL206c/Pgc1:length_22_V	ggttctggatcaggttcaCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTGTTGTCGTACGAtaactccaggagTATATAAACTCAT
3	YPL206c/Pgc1:length_24_V	ggttctggatcaggttcaCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTGTTGTCGTAGTTGTCCGAtaactccaggagTATATAAACTCAT
3	YPL206c/Pgc1:length_26_V	ggttctggatcaggttcaCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTGTTGTCGTAGTTGTCGTAGTTCGAtaactccaggagTATATAAACTCAT
3	YPL206c/Pgc1:length_26_V_shuffled_1	ggttctggatcaggttcaCTTCTATATTCCAAATTCTATTGGGTAGTGGTCTGGCATTTGTTGGCTGTTAAGGTCTCTGTTGTATGCTTTATTGTCCTTGTTGGTATTATTCGAtaactccaggagTATATAAACTCAT
3	YPL206c/Pgc1:length_26_V_shuffled_2	ggttctggatcaggttcaCTTCTATATTCCAAAGTTGTCGGTTGGTTGTTTATTATTTGGCTTGTATATTCTGTTGTCGTGTGCGCTTTCAAGCATGTTGTCGTATTGATTCGAtaactccaggagTATATAAACTCAT
3	YPL206c/Pgc1:length_26_VA	ggttctggatcaggttcaCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTGTTGCTGTCGCAGTAGCCGTTCGAtaactccaggagTATATAAACTCAT
3	YPL206c/Pgc1:TMD_shuffled_1	ggttctggatcaggttcaCTTCTATATTCCAAACTTGGTATTTCTGTCTGGTTGATTTGCGTGTGGTTTTTCATTTATAAGGCTTTGCATCGAACCATTCATTTTtaactccaggagTATATAAACTCAT
3	YPL206c/Pgc1:TMD_shuffled_2	ggttctggatcaggttcaCTTCTATATTCCAAATGCTGGCTTGGTGTGTATTTGATTAAGATTGTCCATTTTATTTCTTGGTTCTTGGCTCGAACCATTCATTTTtaactccaggagTATATAAACTCAT
3	YPL232w/Sso1:length_24_V	ggttctggatcaggttcaAAGAACAAGATTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YPL232w/Sso1:length_26_V	ggttctggatcaggttcaAAGAACAAGATTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCGTTGTCGTAGTTAAAtaactccaggagTATATAAACTCAT
3	YPL232w/Sso1:length_26_V_shuffled_1	ggttctggatcaggttcaAAGAACAAGATTAGAGTCGTCTTGGTTGTCGCCGTAGTTGTTGTACCAATTTGGGTTGTTGCCGTCTTCGTTATCGTAGTTTGTATTGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YPL232w/Sso1:length_26_V_shuffled_2	ggttctggatcaggttcaAAGAACAAGATTAGAGTAGTTGTTATCATTGTCGCCGCCATTCCATTCGTCGTCGTTGTAGTAGTCGTTTGTGTCGTTTTGGTTGTTTGGGTTAAAtaactccaggagTATATAAACTCAT
3	YPL232w/Sso1:length_26_VA	ggttctggatcaggttcaAAGAACAAGATTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCGTTGCTGTCGCAAAAtaactccaggagTATATAAACTCAT
3	YPL232w/Sso1:TMD_shuffled_1	ggttctggatcaggttcaAAGAACAAGATTAGATGTGCCTTGGTTTGGGTCATTGTCCCAGTCGTTATCATTTTCGTAGTTGTAGTTGTTGTCGCCGTTAAAACGCGTtaactccaggagTATATAAACTCAT
3	YPL232w/Sso1:TMD_shuffled_2	ggttctggatcaggttcaAAGAACAAGATTAGAGTCGCCGTAGCCATTTTCGTCCCAATCTGGTGTGTTGTTGTTGTAGTCGTCGTTATTGTTTTGGTTAAAACGCGTtaactccaggagTATATAAACTCAT
3	YPR133w-A/Tom5:length_22_V	ggttctggatcaggttcaAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGGTTGTCGTAGTTAAAtaactccaggagTATATAAACTCAT
3	YPR133w-A/Tom5:length_24_V	ggttctggatcaggttcaAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGGTTGTCGTAGTTGTCGTAAAAtaactccaggagTATATAAACTCAT
3	YPR133w-A/Tom5:length_26_V	ggttctggatcaggttcaAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGGTTGTCGTAGTTGTCGTAGTTGTCAAAtaactccaggagTATATAAACTCAT
3	YPR133w-A/Tom5:length_26_V_shuffled_1	ggttctggatcaggttcaAAAACCTTGAAACAGGTATGGGTCGTCGTGTGGCCAATGGCGGCTATCTTGCATGTACTTTCCTTTGTGGTTGTTGCTTATGTCGTTGCCGTTAAAtaactccaggagTATATAAACTCAT
3	YPR133w-A/Tom5:length_26_V_shuffled_2	ggttctggatcaggttcaAAAACCTTGAAACAGTCCTTGTATCTTGTACATGTTGCTGTCGTTGCCTGGTTTTGGGTGATCCCAGTGGCGGTTGTCGTTGTCATGGCTGTAAAAtaactccaggagTATATAAACTCAT
3	YPR133w-A/Tom5:length_26_VA	ggttctggatcaggttcaAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGGTTGCTGTCGCAGTAGCCGTTGCTAAAtaactccaggagTATATAAACTCAT
3	YPR133w-A/Tom5:TMD_shuffled_1	ggttctggatcaggttcaAAAACCTTGAAACAGGCTGTGTCCGCGATGCCACATATCGTTTGGTTGTGGGCTCTTGTGTTTTATGCCAAAAAGCAATGGAAAtaactccaggagTATATAAACTCAT
3	YPR133w-A/Tom5:TMD_shuffled_2	ggttctggatcaggttcaAAAACCTTGAAACAGTTGGTGGTGGCCTGGTTTATCCTTTGGGCTCCATCCATGTATCATGTTGCGGCTAAAAAGCAATGGAAAtaactccaggagTATATAAACTCAT
3	YPR183w/Dpm1:length_22_V	ggttctggatcaggttcaTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACGTTGTCCATtaactccaggagTATATAAACTCAT
3	YPR183w/Dpm1:length_24_V	ggttctggatcaggttcaTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACGTTGTCGTAGTTCATtaactccaggagTATATAAACTCAT
3	YPR183w/Dpm1:length_26_V	ggttctggatcaggttcaTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACGTTGTCGTAGTTGTCGTACATtaactccaggagTATATAAACTCAT
3	YPR183w/Dpm1:length_26_V_shuffled_1	ggttctggatcaggttcaTTTGGCGCCAATAACACTCTTATTCTACTTGTATTCTTCGTATACGTTATTTACATCCTGGTCTGCCAGTACGTTTGGGTTGTCTTCTTCTCCCATtaactccaggagTATATAAACTCAT
3	YPR183w/Dpm1:length_26_V_shuffled_2	ggttctggatcaggttcaTTTGGCGCCAATAACGTATGCCAGATTGTATTCCTAGTTCTTGTCATCCTGTACTGGTCCTTCTACATTACTGTCGTTCTTTTCTACTTCGTTCATtaactccaggagTATATAAACTCAT
3	YPR183w/Dpm1:length_26_VA	ggttctggatcaggttcaTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACGTTGCTGTCGCAGTAGCCCATtaactccaggagTATATAAACTCAT
3	YPR183w/Dpm1:TMD_shuffled_1	ggttctggatcaggttcaTTTGGCGCCAATAACATCTTCTTCATTTACTACTTCCTACTTATTTCCCAGTTCCTGTACGTTTGCTGGCTTACTCATTTGGTCTTTtaactccaggagTATATAAACTCAT
3	YPR183w/Dpm1:TMD_shuffled_2	ggttctggatcaggttcaTTTGGCGCCAATAACCTGGTTTTCTTCTCCTTCCTATGCATTTACATCCAGCTTTACATTTACTTCTGGCTTACTCATTTGGTCTTTtaactccaggagTATATAAACTCAT
4	0_YDL241w:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCtaactccaggagTATATAAACTCAT
4	1_YDR498c/Sec20:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTtaactccaggagTATATAAACTCAT
4	2_YBL100c:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTtaactccaggagTATATAAACTCAT
4	3_YPR133w-A/Tom5:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAtaactccaggagTATATAAACTCAT
4	4_YER100w/Ubc6:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGAAAtaactccaggagTATATAAACTCAT
4	5_YER120w/Scs2:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACAGAtaactccaggagTATATAAACTCAT
4	6_YOL044w/Pex15:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAtaactccaggagTATATAAACTCAT

4	7_YHL031c/Gos1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACATGGtaactccaggagTATATAAACTCAT
4	8_YLR078c/Bos1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAAtaactccaggagTATATAAACTCAT
4	9_YDR200c/Far9:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGtaactccaggagTATATAAACTCAT
4	10_YOL018c/Tlg2:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAAtaactccaggagTATATAAACTCAT
4	11_YOR106w/Vam3:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAAGTtaactccaggagTATATAAACTCAT
4	12_YDR210w:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTTCTTATTAGAAtaactccaggagTATATAAACTCAT
4	13_YDR498c/Sec20:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTtaactccaggagTATATAAACTCAT
4	14_YBL100c:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTtaactccaggagTATATAAACTCAT
4	15_YDL241w:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCtaactccaggagTATATAAACTCAT
4	16_YOR045w/Tom6:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGACCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAAtaactccaggagTATATAAACTCAT
4	17_YOR075w/Ufe1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGGTtaactccaggagTATATAAACTCAT
4	18_YPL200w/Csm4:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAtaactccaggagTATATAAACTCAT
4	19_YMR161w/Hlj1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTAGTtaactccaggagTATATAAACTCAT
4	20_YDR086c/Sss1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGtaactccaggagTATATAAACTCAT
4	21_YPL206c/Pgc1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAtaactccaggagTATATAAACTCAT
4	22_YNL111c/Cyb5:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACtaactccaggagTATATAAACTCAT
4	23_YDR468c/Tlg1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGCTtaactccaggagTATATAAACTCAT
4	24_YMR197c/Vti1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAtaactccaggagTATATAAACTCAT
4	25_YAL014c/Syn8:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTATTAtaactccaggagTATATAAACTCAT
4	26_YDL012c:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccaggagTATATAAACTCAT
4	27_YBR016w:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	28_YJL119c:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTtaactccaggagTATATAAACTCAT
4	29_YDR034W-B:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACtaactccaggagTATATAAACTCAT
4	30_YPR133w-A/Tom5:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAtaactccaggagTATATAAACTCAT
4	31_YER100w/Ubc6:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGAAAtaactccaggagTATATAAACTCAT
4	32_YER120w/Scs2:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACAGAtaactccaggagTATATAAACTCAT
4	33_YOL044w/Pex15:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAtaactccaggagTATATAAACTCAT
4	34_YHL031c/Gos1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACATGGtaactccaggagTATATAAACTCAT
4	35_YLR078c/Bos1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAAtaactccaggagTATATAAACTCAT
4	36_YDR200c/Far9:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGtaactccaggagTATATAAACTCAT
4	37_YOL018c/Tlg2:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAAtaactccaggagTATATAAACTCAT
4	38_YOR106w/Vam3:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAAGTtaactccaggagTATATAAACTCAT
4	39_YDR210w:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTTCTTATTAGAAtaactccaggagTATATAAACTCAT
4	40_YFL046w/Fmp32:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGtaactccaggagTATATAAACTCAT
4	41_YIL065c/Fis1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAtaactccaggagTATATAAACTCAT
4	42_YER019c-A/Sbh2:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGACTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGtaactccaggagTATATAAACTCAT
4	43_YPR183w/Dpm1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGACTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATtaactccaggagTATATAAACTCAT
4	44_YER087c-B/Sbh1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTtaactccaggagTATATAAACTCAT
4	45_YLR026c/Sed5:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCAATtaactccaggagTATATAAACTCAT
4	46_YKL006c-A/Sft1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGtaactccaggagTATATAAACTCAT
4	47_YOR036w/Pep12:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAAAtaactccaggagTATATAAACTCAT
4	48_YNL188w/Kar1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGGtaactccaggagTATATAAACTCAT
4	49_YBL091c-A/Scs22:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGACTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACTACtaactccaggagTATATAAACTCAT
4	50_YOR045w/Tom6:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAAtaactccaggagTATATAAACTCAT
4	51_YOR075w/Ufe1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGGTtaactccaggagTATATAAACTCAT
4	52_YPL200w/Csm4:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAtaactccaggagTATATAAACTCAT

4	53_YMR161w/Hlj1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTAGTtaactccaggagTATATAAACTCAT
4	54_YDR086c/Sss1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGtaactccaggagTATATAAACTCAT
4	55_YPL206c/Pgc1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAtaactccaggagTATATAAACTCAT
4	56_YNL111c/Cyb5:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACtaactccaggagTATATAAACTCAT
4	57_YDR468c/Tlg1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGCTtaactccaggagTATATAAACTCAT
4	58_YMR197c/Vti1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAtaactccaggagTATATAAACTCAT
4	59_YAL014c/Syn8:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTATTAtaactccaggagTATATAAACTCAT
4	60_YDL012c:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccaggagTATATAAACTCAT
4	61_YBR016w:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	62_YJL119c:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTtaactccaggagTATATAAACTCAT
4	63_YDR034W-B:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACtaactccaggagTATATAAACTCAT
4	64_YDR498c/Sec20:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTtaactccaggagTATATAAACTCAT
4	65_YBL100c:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTTGTtaactccaggagTATATAAACTCAT
4	66_YDL241w:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAGGCCTTCTCATTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCCGTAGAtaactccaggagTATATAAACTCAT
4	67_YBR162w-A/Ysy6:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCCTAtaactccaggagTATATAAACTCAT
4	68_YLR238w/Far10:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCtaactccaggagTATATAAACTCAT
4	69_YLR093c/Nyv1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGTGGtaactccaggagTATATAAACTCAT
4	70_YAL030w/Snc1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTtaactccaggagTATATAAACTCAT
4	71_YFL046w/Fmp32:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGtaactccaggagTATATAAACTCAT
4	72_YIL065c/Fis1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAtaactccaggagTATATAAACTCAT
4	73_YER019c-A/Sbh2:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGtaactccaggagTATATAAACTCAT
4	74_YPR183w/Dpm1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATtaactccaggagTATATAAACTCAT
4	75_YER087c-B/Sbh1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTtaactccaggagTATATAAACTCAT
4	76_YLR026c/Sed5:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCAATtaactccaggagTATATAAACTCAT
4	77_YKL006c-A/Sft1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGtaactccaggagTATATAAACTCAT
4	78_YOR036w/Pep12:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAAAtaactccaggagTATATAAACTCAT
4	79_YNL188w/Kar1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGGtaactccaggagTATATAAACTCAT
4	80_YBL091c-A/Scs22:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACTACtaactccaggagTATATAAACTCAT
4	81_YPR133w-A/Tom5:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAAAtaactccaggagTATATAAACTCAT
4	82_YER100w/Ubc6:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGAAAtaactccaggagTATATAAACTCAT
4	83_YER120w/Scs2:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACAGAtaactccaggagTATATAAACTCAT
4	84_YOL044w/Pex15:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAAAtaactccaggagTATATAAACTCAT
4	85_YHL031c/Gos1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACATGGtaactccaggagTATATAAACTCAT
4	86_YLR078c/Bos1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAAAtaactccaggagTATATAAACTCAT
4	87_YDR200c/Far9:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAAGtaactccaggagTATATAAACTCAT
4	88_YOL018c/Tlg2:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAAAtaactccaggagTATATAAACTCAT
4	89_YOR106w/Vam3:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAAGTtaactccaggagTATATAAACTCAT
4	90_YDR210w:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTTCTTATTAGAAtaactccaggagTATATAAACTCAT
4	91_YDR498c/Sec20:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTAGAtaactccaggagTATATAAACTCAT
4	92_YBL100c:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTCGTAGAtaactccaggagTATATAAACTCAT
4	93_YNL070w/Tom7:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGCAtaactccaggagTATATAAACTCAT
4	94_YNL131w/Tom22:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGACTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGAAtaactccaggagTATATAAACTCAT
4	95_YIL004c/Bet1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTACAtaactccaggagTATATAAACTCAT
4	96_YOR324c/Frt1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGACTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGtaactccaggagTATATAAACTCAT
4	97_YPL232w/Sso1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAtaactccaggagTATATAAACTCAT

4	98_YOR327c/Snc2:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCAGCtaactccaggagTATATAAACTCAT
4	99_YDR281c/Phm6:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGtaactccaggagTATATAAACTCAT
4	100_YBR162w-A/Ysy6:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCCTAtaactccaggagTATATAAACTCAT
4	101_YLR238w/Far10:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCtaactccaggagTATATAAACTCAT
4	102_YLR093c/Nyv1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGTGGtaactccaggagTATATAAACTCAT
4	103_YAL030w/Snc1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTtaactccaggagTATATAAACTCAT
4	104_YOR045w/Tom6:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCAAtaactccaggagTATATAAACTCAT
4	105_YOR075w/Ufe1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAGGTtaactccaggagTATATAAACTCAT
4	106_YPL200w/Csm4:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAGAAtaactccaggagTATATAAACTCAT
4	107_YMR161w/Hlj1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTAGTtaactccaggagTATATAAACTCAT
4	108_YDR086c/Sss1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAAGtaactccaggagTATATAAACTCAT
4	109_YPL206c/Pgc1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGAtaactccaggagTATATAAACTCAT
4	110_YNL111c/Cyb5:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAACtaactccaggagTATATAAACTCAT
4	111_YDR468c/Tlg1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTGCTtaactccaggagTATATAAACTCAT
4	112_YMR197c/Vti1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCTCAtaactccaggagTATATAAACTCAT
4	113_YAL014c/Syn8:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTATTAtaactccaggagTATATAAACTCAT
4	114_YDL012c:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTtaactccaggagTATATAAACTCAT
4	115_YBR016w:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	116_YJL119c:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTTTTtaactccaggagTATATAAACTCAT
4	117_YDR034W-B:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTGACtaactccaggagTATATAAACTCAT
4	118_YPR133w-A/Tom5:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGCGTAGAtaactccaggagTATATAAACTCAT
4	119_YER100w/Ubc6:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCCTAATGATAGTTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGCGTAGAtaactccaggagTATATAAACTCAT
4	120_YER120w/Scs2:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAATGAATCATCCAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACCGTAGAtaactccaggagTATATAAACTCAT
4	121_YOL044w/Pex15:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGCGTAGAtaactccaggagTATATAAACTCAT
4	122_YHL031c/Gos1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAGAAGGAAGAAAAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACACGTAGAtaactccaggagTATATAAACTCAT
4	123_YLR078c/Bos1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGCGTAGAtaactccaggagTATATAAACTCAT
4	124_YDR200c/Far9:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTCGTAGAtaactccaggagTATATAAACTCAT
4	125_YOL018c/Tlg2:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGCGTAGAtaactccaggagTATATAAACTCAT
4	126_YOR106w/Vam3:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAACAAATGCGGTAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTACGTAGAtaactccaggagTATATAAACTCAT
4	127_YDR210w:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTTCTTATTACGTAGAtaactccaggagTATATAAACTCAT
4	128_YDL241w:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAGGCCTTCTCATTCTCTCGCCAAATTTACCATACTGTTCATGCTAGTGTTGTATACAATTGTCAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	129_YbR056W-A/Mnc1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCTAATTtaactccaggagTATATAAACTCAT
4	130_YNL070w/Tom7:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGCAtaactccaggagTATATAAACTCAT
4	131_YNL131w/Tom22:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGAAtaactccaggagTATATAAACTCAT
4	132_YIL004c/Bet1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTACAtaactccaggagTATATAAACTCAT
4	133_YOR324c/Frt1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGtaactccaggagTATATAAACTCAT
4	134_YPL232w/Sso1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAtaactccaggagTATATAAACTCAT
4	135_YOR327c/Snc2:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCAGCtaactccaggagTATATAAACTCAT
4	136_YDR281c/Phm6:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGCTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGtaactccaggagTATATAAACTCAT
4	137_YFL046w/Fmp32:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGGtaactccaggagTATATAAACTCAT
4	138_YIL065c/Fis1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGAtaactccaggagTATATAAACTCAT
4	139_YER019c-A/Sbh2:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGACGtaactccaggagTATATAAACTCAT
4	140_YPR183w/Dpm1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCATtaactccaggagTATATAAACTCAT
4	141_YER087c-B/Sbh1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTtaactccaggagTATATAAACTCAT
4	142_YLR026c/Sed5:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCAATtaactccaggagTATATAAACTCAT

4	143_YKL006c-A/Sft1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAAGtaactccaggagTATATAAACTCAT
4	144_YOR036w/Pep12:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAAAtaactccaggagTATATAAACTCAT
4	145_YNL188w/Kar1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGGtaactccaggagTATATAAACTCAT
4	146_YBL091c-A/Scs22:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACTACtaactccaggagTATATAAACTCAT
4	147_YbR056W-A/Mnc1:TAvariants_1_1_	ggttctggatcaggttcaGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCTAATTtaactccaggagTATATAAACTCAT
4	148_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATCTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	149_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATGTCTGGCTGCATTATCTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	150_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATGTCTGGCTGCATTATGTATATCTTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	151_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	152_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATCTCTGGCTGCATTATCTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	153_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATCTCTGGCTGCATTATGTATATCTTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	154_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATCTCTGGCTGCATTATGTATATGCTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	155_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATCTATATCTTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	156_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATCTATATGCTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	157_YBR016w:C_to_S_pair	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATCTTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	158_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATCTCTGGCTGCATTATCTATATGCTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	159_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATCTCTGGCTGCATTATGTATATCTTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	160_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATCTCTGGCTGCATTATGTATATGCTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	161_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATGTCTGGCTGCATTATCTATATCTTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	162_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATGTCTGGCTGCATTATCTATATGCTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	163_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATGTCTGGCTGCATTATGTATATCTTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	164_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATCTCTGGCTGCATTATCTATATCTTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	165_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATCTCTGGCTGCATTATCTATATGCTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	166_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATCTCTGGCTGCATTATGTATATCTTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	167_YBR016w:C_to_S_triplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATCTATATCTTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	168_YBR016w:C_to_S_quadruplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATCTCTGGCTGCATTATCTATATCTTGCACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	169_YBR016w:C_to_S_quadruplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATCTCTGGCTGCATTATCTATATGCTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	170_YBR016w:C_to_S_quadruplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATCTCTGGCTGCATTATGTATATCTTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	171_YBR016w:C_to_S_quadruplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTCTCTGGCTGCATGTCTGGCTGCATTATCTATATCTTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	172_YBR016w:C_to_S_quadruplet	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATCTCTGGCTGCATTATCTATATCTTCTACCATGGATATGCTATTCtaactccaggagTATATAAACTCAT
4	173_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAAGCACTGTGCCTGACCTGTTGTTGTATGGCTGCTTTAGGTCTAATGGCAGATTGCATATTCtaactccaggagTATATAAACTCAT
4	174_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAAATGGCATGTCTGATATTACTACTGACCTGCTGTATGGGTGCTGCAGCTTGCGATTGTTTCtaactccaggagTATATAAACTCAT
4	175_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAATGCTGCATGTGTCTGATGTGTCTAGCAGCAGCTCTGGATGCTTGTGGTTTAACCATATTCtaactccaggagTATATAAACTCAT
4	176_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAATGTATGGCACTGTTAATAGGTTGCACCCTAGCTCTGGCTTGTATGGCATGCTGTGATTTCtaactccaggagTATATAAACTCAT
4	177_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAAATGGCTATAATGTGTCTGCTGTTAGATGGTGCAGCTGCATGTTGTTGCTGCCTAACCTTCtaactccaggagTATATAAACTCAT
4	178_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAACTAACCCTGTTATGTTGTATACTGGCAGGTGCTGCAATGATGTGTGCTGATTGCTGCTTCtaactccaggagTATATAAACTCAT
4	179_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAAGCATGCATGATATGTGCTCTGTTAGATGGTATGCTAGCAGCTTGTACCTGTTGCCTGTTCtaactccaggagTATATAAACTCAT
4	180_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAAGATACCGCTTGCTGTTGTGCTTGCGGTATGCTGCTAATATTAATGGCATGTCTGGCATTCtaactccaggagTATATAAACTCAT
4	181_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAATGCGATATGTGTGCTATGGCTTGTATATGTGGTGCACTGTTACTATGCCTGACCGCATTCtaactccaggagTATATAAACTCAT
4	182_YBR016w:shuffled	ggttctggatcaggttcaCAGAGGGGTAACGAAGCTGCATGTTGCTTAGATGGTCTGGCTTGCTGTTGTATGCTGATACTAACCGCAATGTTCtaactccaggagTATATAAACTCAT
4	183_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAATTAGCGACTCTATGCGGCATGGCGCTAGATTTATGTTTGGACTGTCTGTGTTGTGGTTTTtaactccaggagTATATAAACTCAT
4	184_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAAACTCTAATGTGTGGTTGTTTGTTATGTGACCTGTGTGGCGATGCGTGCGCGTTACTATTTtaactccaggagTATATAAACTCAT
4	185_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAACTAGGTCTATGTTGTGCGTGTGCGTTAATGGACTGCTGTTTGCTGGATACTTTAGGCTTTtaactccaggagTATATAAACTCAT
4	186_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAAACTTTACTGTGTTGTGGTGATTGTGCGTGTTTACTACTATGCGCGATGGGCTTGGACTTTtaactccaggagTATATAAACTCAT
4	187_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAACTAATGCTGTTATTATGTTGTCTAGGTACTGCGTGTTTGTGCGGCGATGCGTGTGACTTTtaactccaggagTATATAAACTCAT

4	188_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTTGTTATGCTGTTGTGGCGCGGATCTAGGTCTGCTATTAACTTGTGCGTGTATGTTTtaactccagg agTATATAAACTCAT
4	189_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAATGCGGCTTGTGTGCGTGTATGCTAACTCTATGTGCGTTACTGTGTGACGATGGTTTATTTtaactccagg agTATATAAACTCAT
4	190_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAAATGTTGTGTGACTGCCTATGTCTGCTAGGTTGTTTAGGCTTATGTGATACTGCGGCGTTTtaactccagg agTATATAAACTCAT
4	191_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAAGCGTTAATGGATGGTTGTTGTACTGCGGACCTAGGCCTATTGTGTCTGTGTTGCTTATTTtaactccagg agTATATAAACTCAT
4	192_YDL012c:shuffled	ggttctggatcaggttcaAGTTCTGGAAATGAAACTTTGATGTGTCTATGTGCGGACTGTGATTTACTGTGTGCGTGCGGTTTAGGCCTATTTtaactccagg agTATATAAACTCAT
4	193_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAAGCTTGTGCATTAATGTGTGATCTGACCCTGTGTGCATGCTGCATGCTAGCTGGTATATTCtaactccag gagTATATAAACTCAT
4	194_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAATTATGTCTAGCTCTGTGTCTGATGATAGGTTGTACCTGCTGCATGGCAGCAGCTGATTTCtaactccag gagTATATAAACTCAT
4	195_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAAACCTGTATGCTGCTATGTGCTGCAATAGCATGTGATTGCTGCTTACTGGGTATGGCTTTCtaactccag gagTATATAAACTCAT
4	196_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAACTGTGTGATCTGATGTGTATGGCTGCAACCTGTCTATGCTGCGCAATAGGTTTAGCTTTCtaactccag gagTATATAAACTCAT
4	197_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAACTGTGTGCAATGGCATGTTTAATAATGGATTGTCTGTGCTGCCTAGCTGCTACCGGTTTCtaactccag gagTATATAAACTCAT
4	198_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAAATGTGTGCAGCTGCTTGTGATCTGATGATATGTGCATGCTGCTTAGGTACCCTACTGTTCtaactccag gagTATATAAACTCAT
4	199_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAACTATGTGCACTGATGTGTTTAACCATGGCATGTCTGTGCTGCATAGCTGCTGATGGTTTCtaactccag gagTATATAAACTCAT
4	200_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAAGCTTGTATGTTAATATGTGCACTAATGGATTGTGCATGCTGCACCGGTGCTCTGCTGTTCtaactccag gagTATATAAACTCAT
4	201_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAAATGTGTATGGCTCTATGTTTACTGGATCTGTGTGCTTGCTGCGCAGGTACCGCAATATTCtaactccag gagTATATAAACTCAT
4	202_YBR016w:shuffled_fixed_C	ggttctggatcaggttcaCAGAGGGGTAACGAAATGTGTATGGCACTGTGTATATTAGCTGGTTGTCTGTGCTGCGCTGCACTAACCGATTTCtaactccag gagTATATAAACTCAT
4	203_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAATTGTGTGGTTTAATGTGTCTAGACCTAGCGTGTCTGTGCTGTGGCGCGTTAGATACTTTTtaactccagg agTATATAAACTCAT
4	204_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAAGCGTGTCTAATGGACTGTGGTTTGCTGGCGTGTCTATGCTGTGATACTTTATTAGGCTTTtaactccagg agTATATAAACTCAT
4	205_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAATTATGTCTATTGCTGTGTTTACTAGATGCGTGTATGTGCTGTGGTACTGACGCGGGCTTTtaactccagg agTATATAAACTCAT
4	206_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAAGGCTGTGATCTACTGTGTGCGGCGACTCTATGTATGTGCTGTGACTTGTTATTAGGTTTTtaactccagg agTATATAAACTCAT
4	207_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAACTGTGTCTAGGCTTATGTCTAATGGCGTTGTGTTTATGCTGTGACGGTACTGATGCGTTTtaactccagg agTATATAAACTCAT
4	208_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAATTATGTCTGACTGGTTGTGCGGCGGACGATTGTATGTGCTGTTTAGGCTTGCTACTATTTtaactccagg agTATATAAACTCAT
4	209_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAATTATGTGGTCTGCTATGTGCGACTGACCTATGTGATTGCTGTGGCTTGATGGCGTTATTTtaactccagg agTATATAAACTCAT
4	210_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAACTATGTGATGGCCTATGTGCGATGCTGACTTGTTTATGCTGTGCGTTGTTAGGTGACTTTtaactccagg agTATATAAACTCAT
4	211_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAACTATGTGGTTTAGACTGTCTGCTATTAGATTGTGGCTGCTGTATGACTTTGGCGGCGTTTtaactccagg agTATATAAACTCAT
4	212_YDL012c:shuffled_fixed_C	ggttctggatcaggttcaAGTTCTGGAAATGAAGGTTGTCTACTGATGTGTCTAGCGTTAGACTGTGCGTGCTGTGGCGATTTATTGACTTTTtaactccagg agTATATAAACTCAT
4	213_YOR045w/Tom6:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTCGTAGAtaactcc aggagTATATAAACTCAT
4	214_YOR075w/Ufe1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGGAAGAACTGCTAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTACGTAGAtaactcc aggagTATATAAACTCAT
4	215_YPL200w/Csm4:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATACGTAGAtaactcca ggagTATATAAACTCAT
4	216_YMR161w/Hlj1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaTCGCAATTAAAAAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTCGTAGAtaactccag gagTATATAAACTCAT
4	217_YDR086c/Sss1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCCGTAGAtaactc caggagTATATAAACTCAT
4	218_YPL206c/Pgc1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTCGTAGAtaactcca ggagTATATAAACTCAT
4	219_YNL111c/Cyb5:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGCGTAGAtaactc caggagTATATAAACTCAT
4	220_YDR468c/Tlg1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGAAAAATACGACGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTCGTAGAtaactcc aggagTATATAAACTCAT
4	221_YMR197c/Vti1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCCGTAGAtaactcca ggagTATATAAACTCAT
4	222_YAL014c/Syn8:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAGGTTCAAGGATAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTACGTAGAtaactc caggagTATATAAACTCAT
4	223_YBR016w:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTACGTAGAtaact ccaggagTATATAAACTCAT
4	224_YJL119c:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTCGTAGAtaactcc aggagTATATAAACTCAT
4	225_YDR034W-B:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTCGTAGAtaactcc aggagTATATAAACTCAT
4	226_YDR498c/Sec20:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGAGACGTAGGCGTtaactc caggagTATATAAACTCAT
4	227_YBL100c:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGATAGAGAACGAATAGCAGCAGTACGAGCCAACATTATTATATGTGCGTGTTTTTTTTATTTATTTAGACGTAGGCGTtaactc caggagTATATAAACTCAT
4	228_YPL192c/Prm3:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAtaactcc aggagTATATAAACTCAT
4	229_YMR183c/Sso2:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTGGAAtaactcca ggagTATATAAACTCAT
4	230_YbR056W-A/Mnc1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCTAATTtaactcc aggagTATATAAACTCAT
4	231_YBR162w-A/Ysy6:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCCTAtaactcc aggagTATATAAACTCAT
4	232_YLR238w/Far10:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTTCCtaactcca ggagTATATAAACTCAT

4	233_YLR093c/Nyv1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGTGGtaactccaggagTATATAAACTCAT
4	234_YAL030w/Snc1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGTtaactccaggagTATATAAACTCAT
4	235_YFL046w/Fmp32:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGCGTAGAtaactccaggagTATATAAACTCAT
4	236_YIL065c/Fis1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTACGTAGAtaactccaggagTATATAAACTCAT
4	237_YER019c-A/Sbh2:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGCGTAGAtaactccaggagTATATAAACTCAT
4	238_YPR183w/Dpm1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACCGTAGAtaactccaggagTATATAAACTCAT
4	239_YER087c-B/Sbh1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTCGTAGAtaactccaggagTATATAAACTCAT
4	240_YLR026c/Sed5:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaATAAAGAGTAATAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCCGTAGAtaactccaggagTATATAAACTCAT
4	241_YKL006c-A/Sft1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTCGTAGAtaactccaggagTATATAAACTCAT
4	242_YOR036w/Pep12:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGCGTAGAtaactccaggagTATATAAACTCAT
4	243_YNL188w/Kar1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATCGTAGAtaactccaggagTATATAAACTCAT
4	244_YBL091c-A/Scs22:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGGTAACGGGCAAAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACCGTAGAtaactccaggagTATATAAACTCAT
4	245_YDL012c:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTCGTAGAtaactccaggagTATATAAACTCAT
4	246_YPR133w-A/Tom5:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAAAACCTTGAAACAGGCCGCTTATGTGGCTGCGTTTCTTTGGGTTTCCCCAATGATCTGGCATTTGGTGAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	247_YER100w/Ubc6:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCCTAATGATAGTTCTTCAATGGTTTATATTGGTATCGCTATTTTTTTGTTTTTGGTTGGCCTTTTTATGAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	248_YER120w/Scs2:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAATGAATCATCCAGCATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	249_YOL044w/Pex15:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCATTTCACAAGAAGTGTGCTGAACAAAAACGGACTTCTGCTCACAGGTCTGCTGCTACTCTTATGTTTGAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	250_YHL031c/Gos1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAGAAGGAAGAAAAACGCGTTTGTATTGGCCACGATAACCACCCTTTGTATACTGTTTTTGTTTTTCACAAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	251_YLR078c/Bos1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGTGTTCAAAGATAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	252_YDR200c/Far9:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaACAAGAGTTAGTAAGGGAATGCTCTTTGGAGTAGTTGCCATTTCATTCGGTTTAGTTGCGACGGCCGTTAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	253_YOL018c/Tlg2:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaACTCAAAAATGTAAAGTCATTTTATTACTAACATTGTGCGTAATAGCGCTCTTTTTCTTTGTTATGTTGAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	254_YOR106w/Vam3:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAACAAATGCGGTAAGGTCACCCTAATCATTATAATAGTTGTGTGCATGGTGGTATTGCTTGCTGTATTAAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	255_YDR210w:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGAAAGCTGCTTGGACAGTTGTTTAAAATGTTTGTGTTGTTGTTTCTTATTAAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	256_YGL098w/Use1:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGtaactccaggagTATATAAACTCAT
4	257_YPL192c/Prm3:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAAAtaactccaggagTATATAAACTCAT
4	258_YMR183c/Sso2:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTGGAAtaactccaggagTATATAAACTCAT
4	259_YNL070w/Tom7:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGGCAtaactccaggagTATATAAACTCAT
4	260_YNL131w/Tom22:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCGAAtaactccaggagTATATAAACTCAT
4	261_YIL004c/Bet1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTACAtaactccaggagTATATAAACTCAT
4	262_YOR324c/Frt1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAAGtaactccaggagTATATAAACTCAT
4	263_YPL232w/Sso1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAAAtaactccaggagTATATAAACTCAT
4	264_YOR327c/Snc2:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCAGCtaactccaggagTATATAAACTCAT
4	265_YDR281c/Phm6:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGATGtaactccaggagTATATAAACTCAT
4	266_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTTGTTTATGTTGTCTATGCTGTTGTCTACTATGTACCTTCTGTAACTGTCACATTAGAGACGTTTTTtaactccaggagTATATAAACTCAT
4	267_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTTGTTTCATTTGTCTATGTAACTGTTGTTTAACCAGACACCTATGCTGTCTATGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT
4	268_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTTTCTGTCTATGTTTATGTTGTTGTTGTCTACTAATTCACTGCAGAAACACCTGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT
4	269_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTACCTGTTGTTGTAACTTAAGATGTCTAATTTGTCTATGTTTCTGTTGCCTATGTCACGACGTTTTTtaactccaggagTATATAAACTCAT
4	270_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTCTAATTTGTTGTTGTCACACCCTATGTTTATGCCTATGTTGTTGTTTCTGTAGAAACGACGTTTTTtaactccaggagTATATAAACTCAT
4	271_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTTTACACTGTTGCCTATGTTGTTGTATTTGTTGTTGTCTACTATGTACCAGAAACTTCGACGTTTTTtaactccaggagTATATAAACTCAT
4	272_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTAACTGTTGTTGTCACCTATGTATTTGTTGTTGTTTCACCTGCCTACTATTAAGATGTGACGTTTTTtaactccaggagTATATAAACTCAT
4	273_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTTGTCTAAACTGTTGTTTACTATTCTGCTGTATTCACACCCTATGTTGTTGTAGATGTGACGTTTTTtaactccaggagTATATAAACTCAT
4	274_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTTTCCTAACCCACCTATGTCTATGTTGTTGTTGTTGTATTAACTGTTTAAGATGCGACGTTTTTtaactccaggagTATATAAACTCAT
4	275_YDR034W-B:shuffled	ggttctggatcaggttcaAGAGAATCCGGTGGTCACTTAAACATTTGTCTATGTAGACTATGTTGCTGTCTATTCACCTGTTGTTGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT
4	276_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAACTTCTGTTGTATTCTACACTGTTGCCTATGTCTATTAAGAACCTGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT
4	277_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTCTAATTTGTTGTCTACTATTCTGTTGCAACTGTAGAACCTTACACTGTTGTGACGTTTTTtaactccaggagTATATAAACTCAT

4	278_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTATTAGATGTTGTCTAACCCACTGTTGCTTCTGTCTAAACTTACTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
4	279_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTATTAACTGTTGTCTACTATTATGTTGCACCTGTTTCCTAAGACACTGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
4	280_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAACCACTGTTGTCTAATTTTATGTTGCTTCTGTAGACTAACCCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
4	281_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAACCACTGTTGTTTACTAAGATGTTGCCTATGTTTCCTAACCATTTGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
4	282_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTCTATTCTGTTGTACCAACCACTGTTGCATTTGTCTATTAAGACTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
4	283_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTCTAATTTGTTGTAGACTATTATGTTGCACCTGTTTCAACCACCTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
4	284_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTCTACTATGTTGTCTATTAACCTGTTGCCACTGTATTAACTTCAGATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
4	285_YDR034W-B:shuffled_fixed_C	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTATTCACTGTTGTCTATTCAGATGTTGCAACTGTACCCTACTATTATGTTGTGACGTTTTTtaac tccaggagTATATAAACTCAT
4	286_YBR162w-A/Ysy6:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGTGATATCCAAAACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCCGTAG AtaactccaggagTATATAAACTCAT
4	287_YLR238w/Far10:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTCGTAGA taactccaggagTATATAAACTCAT
4	288_YLR093c/Nyv1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCAGAAGGTCAAAAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGCGTAGAt aactccaggagTATATAAACTCAT
4	289_YAL030w/Snc1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCTAAAAATGAAGATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTCGTAGA taactccaggagTATATAAACTCAT
4	290_YOR045w/Tom6:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGCTTTCAAGGAATCTCCACTATACACAATTGCACTAAACGGTGCCTTCTTCGTTGCAGGTGTTGCGTTTATTAGACGTAGGCG TtaactccaggagTATATAAACTCAT
4	291_YOR075w/Ufe1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGGAAGAACTGCTAAAATGACCACTTATGGTGCCATTATCATGGGTGTTTTTATATTGTTCCTAGATTATGTAAGACGTAGGC GTtaactccaggagTATATAAACTCAT
4	292_YPL200w/Csm4:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaTTCGTCATCGCGGAACTAAATTCATTGATAATAGTTTTTTTTATTTCTTTAGTATTTTTGTGGGTATCGATAAGACGTAGGCGT taactccaggagTATATAAACTCAT
4	293_YMR161w/Hlj1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaTCGCAATTAAAAAATATGCTCGTTCTTTTCATCATCTTTATTGTTCTTCCTATGATTAAAGATTACCTGTTTAGACGTAGGCGTt aactccaggagTATATAAACTCAT
4	294_YDR086c/Sss1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAAGGAATACACCAAGATTGTCAAGGCTGTTGGTATTGGTTTTATTGCAGTCGGTATCATTGGTTACGCCATCAGACGTAGGC GTtaactccaggagTATATAAACTCAT
4	295_YPL206c/Pgc1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCTTCTATATTCCAAATGGGTCCATATTAAGTTGTGCGGTTGGTCTATTGCTTATGTGATTTTTTTGTTCCTTAGACGTAGGCGT taactccaggagTATATAAACTCAT
4	296_YNL111c/Cyb5:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCAAAGTAAAGGTAGTGGTACATTGGTTGTCATATTGGCCATTTTAATGCTAGGTGTTGCTTATTATTTGTTGAGACGTAGGC GTtaactccaggagTATATAAACTCAT
4	297_YDR468c/Tlg1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGAAAAATACGACGATTGTTGTATAGGACTTCTTATTGTCGTCTTGATAGTTTTATTAGTTTTGGCATTCATTAGACGTAGGCG TtaactccaggagTATATAAACTCAT
4	298_YMR197c/Vti1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCTAGTTGCTAATAAATTCATAAGCTATGCCATTATCGCAGTCCTTATATTATTGATTTTGCTAGTTTTGTTCAGACGTAGGCGT taactccaggagTATATAAACTCAT
4	299_YAL014c/Syn8:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAGGTTCAAGGATAACGGCAACTGCGTGATTATTCTTGTGCTGATAGTGGTGCTCCTGTTGCTTCTATTAGTAAGACGTAGGC GTtaactccaggagTATATAAACTCAT
4	300_YBR016w:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCAGAGGGGTAACGAAGGTTGTCTGGCTGCATGTCTGGCTGCATTATGTATATGCTGCACCATGGATATGCTAAGACGTAGG CGTtaactccaggagTATATAAACTCAT
4	301_YJL119c:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaACTGAAAAGCATGAAGCCATCACTGTTTTTTTCGTGGTCTTTCTATATATAAACAAAGCCTCTGTAAAGTTTAGACGTAGGCG TtaactccaggagTATATAAACTCAT
4	302_YDR034W-B:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAGAGAATCCGGTGGTTGCTGTAGAACCTGTTGTCACTTCCTATGTTGTCTATGTTTAATTAACCTATGTTGTAGACGTAGGCG TtaactccaggagTATATAAACTCAT
4	303_YLR268w/Sec22:TAvariants_AGA_0_1_	ggttctggatcaggttcaAGATTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCCTCta actccaggagTATATAAACTCAT
4	304_YGL098w/Use1:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCCCGta actccaggagTATATAAACTCAT
4	305_YbR056W-A/Mnc1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCTAAT TtaactccaggagTATATAAACTCAT
4	306_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGTTATGTTTCGAGTTATGTGAATTAAAAAGTGACTGTTGCAGCTTGTGTTTGTGTGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	307_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGGACTTGTGTAGTTGTTTATGCTGTGAGGAAAGCTTCTTATTATTGTGTTGTAAAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	308_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGTGCTGTTGTTTGAGTTTATGTTGTTTCTGTTTATTGAGCGAGGACAAAGAATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	309_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGTGTTTGTGTAAATTATTCTTGTGTTGTAGTTGTGACTGCAGCTTAGAGGAATTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	310_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGAGTTGTGAGGAATGCTGTTTGTTCTGTTGTTGTAGCTTAAAATTGTTAGACTTAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	311_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGTGTTTATGTTTATTAGACTGCTGTTTCGAAGAGAGCAAATTGAGTTGTTTGTGTGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	312_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGTGTTGTTTATGTAGTGAATTGTGTTTCTGCAGCTGTTTAGAGGACTTGTTAAAAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	313_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGTTCTTGTGCAAATGTGAAAGTAGCGAGTTATTATTATGTTGTTTGGACTGTTGTGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	314_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGTTGTGTTGTTTGTGCGAGAAATGTTTCTTATTAAGTGACGAATGTTTAAGCTGTGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	315_YDR210w:shuffled	ggttctggatcaggttcaCAAGGTCAACCAAAGAAATTATTATGTAGTTTAAGCTGTTGCGAGTTGGAATGTTTGGACTGTTTCTGTGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	316_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGAGCTTAGAGTGCAGTTTGTTATGTTTGTTATGTAAATGTTGTTGCGACGAATTCGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	317_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGTTGTTAGAGTGCAGTTTCAGCTGTAAAGACTGTTTATGTTGTTGCTTATTGGAAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	318_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGACTTGTGCAGTTTATTATGTTTGTTATGTAGCTGTTGTTGCTTCGAAAAAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	319_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGTTAGACTGCAGCGAAAAATGTTTCTTATGTTTATGTTGTTGCTTGTTGAGTGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	320_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGAAATTATTGTGCGAAAGTGACTGTTTAAGCTGTTTCTGTTGTTGCTTGTTAGAGGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	321_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGTTGTTCTGCTTATTATTATGTAAAAGCTGTTTGTGTTGTTGCGACAGTGAAGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	322_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGTTGTTATTCTGCTTAAGTTTATGTTTGGACTGTGAGTGTTGTTGCGAAAAAAGCGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT

4	323_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGTTGGAGTTATGCAGCTTAAGTTGTTTATTGTGTGAATGTTGTTGCAAAGACTTCGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	324_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGAAAGAATGCGACTTATTGTGTAGCTTATGTTTCTGTTGTTGCTTAAGTTTGGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	325_YDR210w:shuffled_fixed_C	ggttctggatcaggttcaCAAGGTCAACCAAAGGAGGACTTGTGCAGTTTGGAATGTTTATTATGTAAATGTTGTTGCAGCTTATTCGAATTGGTATGCG ATtaactccaggagTATATAAACTCAT
4	326_YNL070w/Tom7:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGAACGTATTTCCAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGCGT AGAtaactccaggagTATATAAACTCAT
4	327_YNL131w/Tom22:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaACAAAATCCGGAAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCCGT AGAtaactccaggagTATATAAACTCAT
4	328_YIL004c/Bet1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGCCAGAAGATCTGGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTC GTAGAtaactccaggagTATATAAACTCAT
4	329_YOR324c/Frt1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaACGAAATCCATCTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCCGT AGAtaactccaggagTATATAAACTCAT
4	330_YPL232w/Sso1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAAGAACAAGATTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCCG TAGAtaactccaggagTATATAAACTCAT
4	331_YOR327c/Snc2:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGATCTAAAAATGAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCCGT AGAtaactccaggagTATATAAACTCAT
4	332_YDR281c/Phm6:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGGGTTTCTCGAGGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGCGT AGAtaactccaggagTATATAAACTCAT
4	333_YFL046w/Fmp32:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaTCTGTCAAGACTCAAGTCATGCAATGGCTGATAGGTGTCTGTACAGGTACATTTGCACTTGTATTAGCATATATGAGACGTA GGCGTtaactccaggagTATATAAACTCAT
4	334_YIL065c/Fis1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAAGGAAACACTCAAGGGTGTTGTCGTCGCTGGAGGCGTACTAGCCGGCGCTGTGGCCGTGGCTAGTTTCTTCTTAAGACGT AGGCGTtaactccaggagTATATAAACTCAT
4	335_YER019c-A/Sbh2:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaTTCAGAGTCGACTCCCTTGTCGTATTGTTCCTATCTGTCGGTTTCATCTTCTCCGTGATTGCTTTGCATCTATTGAGACGTAGG CGTtaactccaggagTATATAAACTCAT
4	336_YPR183w/Dpm1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaTTTGGCGCCAATAACCTTATCCTTTTCATTACTTTCTGGTCCATTCTGTTCTTCTACGTTTGCTACCAGCTATACAGACGTAGGC GTtaactccaggagTATATAAACTCAT
4	337_YER087c-B/Sbh1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCTAAGAGTAGATCCCTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTAGACGTAGG CGTtaactccaggagTATATAAACTCAT
4	338_YLR026c/Sed5:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaATAAAGAGTAATAGATGGTTAGCCGCAAAGGTTTTTTTTATAATCTTTGTATTTTTCGTTATTTGGGTTTTAGTCAGACGTAG GCGTtaactccaggagTATATAAACTCAT
4	339_YKL006c-A/Sft1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaTTAAAGGCTGGTAATAGCATATGGAGAATGGTGGGTTTGGCGTTATTAATATTCTTCATTCTATATACCCTGTTTAGACGTAG GCGTtaactccaggagTATATAAACTCAT
4	340_YOR036w/Pep12:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAAACGTACGAGCAGATGGAGGGTGTATTTGTTGATTGTGCTTCTCGTAATGCTTCTTTTTATTTTTCTCATTATGAGACGTAG GCGTtaactccaggagTATATAAACTCAT
4	341_YNL188w/Kar1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAAAAAGTATAGGGAATATTTCTTATGGACAATTTGTATTTTAATATTGTTATATTGCAATATATATGTGTATTATAGACGTAG GCGTtaactccaggagTATATAAACTCAT
4	342_YBL091c-A/Scs22:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGGTAACGGGCAAAGTCTGAGTTCCAGAGCATTGCTTATCATCACCGTTATCGCATTGCTCGTCGGCTGGATATACAGACGTA GGCGTtaactccaggagTATATAAACTCAT
4	343_YDL012c:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAGTTCTGGAAATGAAGACTGTCTAGCGGGCTGTCTAGCGGGTTTATGTTTATGCTGTACTTTGGATATGCTGTTTAGACGTA GGCGTtaactccaggagTATATAAACTCAT
4	344_YLR268w/Sec22:TAvariants_AGAAGG_0_1_	ggttctggatcaggttcaAGAAGGTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCC TCtaactccaggagTATATAAACTCAT
4	345_YPL192c/Prm3:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGT TAAAtaactccaggagTATATAAACTCAT
4	346_YMR183c/Sso2:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTG GAAtaactccaggagTATATAAACTCAT
4	347_YbR056W-A/Mnc1:TAvariants_1_5_	ggttctggatcaggttcaGATTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCTAATTGATCTATGT TGTtaactccaggagTATATAAACTCAT
4	348_YbR056W-A/Mnc1:TAvariants_5_1_	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT AATTtaactccaggagTATATAAACTCAT
4	349_YbR056W-A/Mnc1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCT ACGTAGAtaactccaggagTATATAAACTCAT
4	350_YBR162w-A/Ysy6:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGTGATATCCAAAACTTGGTTGGGTATTCTTCTGTTTCTTCTCGTAGGTGGTGGTGTTTTGCAACTAATCAGCTATATCAGACG TAGGCGTtaactccaggagTATATAAACTCAT
4	351_YLR238w/Far10:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaACGCTATGTAATCATTTCACACTATTAACATTTGGAACTATTTCCATCGGGATTATAGCTATTGTCTTCAAGATCCTTAGACGT AGGCGTtaactccaggagTATATAAACTCAT
4	352_YLR093c/Nyv1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCAGAAGGTCAAAAATATTACGTTATTAACTTTCACTATTATACTATTTGTAAGTGCTGCTTTCATGTTTTTCTATCTGAGACGT AGGCGTtaactccaggagTATATAAACTCAT
4	353_YAL030w/Snc1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCTAAAAATGAAGATGTGTCTGGCTTTAGTAATCATCATATTGCTTGTTGTAATCATCGTCCCCATTGCTGTTCACTTTAGACGT AGGCGTtaactccaggagTATATAAACTCAT
4	354_YGL098w/Use1:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTAT TCCCGtaactccaggagTATATAAACTCAT
4	355_YPL192c/Prm3:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGAGAATAAGGGTTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGC AGTTCGTAGAtaactccaggagTATATAAACTCAT
4	356_YMR183c/Sso2:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGCAAGAAAAAACAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTT GTGCGTAGAtaactccaggagTATATAAACTCAT
4	357_YNL070w/Tom7:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGAACGTATTTCCAAAATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTTGGGCTGGAG ACGTAGGCGTtaactccaggagTATATAAACTCAT
4	358_YNL131w/Tom22:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaACAAAATCCGGAAACCTTGCTTGGACTTTGACCACCACTGCTTTGTTACTCGGTGTGCCACTATCCTTATCTATACTTGCCAGA CGTAGGCGTtaactccaggagTATATAAACTCAT
4	359_YIL004c/Bet1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGCCAGAAGATCTGGGATCAGTATAAAAACATGGTTAATAATATTTTTTATGGTAGGCGTGCTATTTTTTTGGGTATGGATTA GACGTAGGCGTtaactccaggagTATATAAACTCAT
4	360_YOR324c/Frt1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaACGAAATCCATCTACCTGTATAGGTATTTCATCATTGACATTATTGCATTTCTTTTGATGGGCGGCTTTATCGTTTACGTCAGA CGTAGGCGTtaactccaggagTATATAAACTCAT
4	361_YPL232w/Sso1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAAGAACAAGATTAGATGTTGGTTGATTGTATTCGCCATCATTGTAGTCGTTGTTGTTGTCGTTGTTGTCCCAGCCGTTGTCAG ACGTAGGCGTtaactccaggagTATATAAACTCAT
4	362_YOR327c/Snc2:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGATCTAAAAATGAGAATGTGTTTATTCTTAGTTGTTATTATTTTACTAGTGGTAATTATCGTTCCTATCGTCGTCCATTTCAGA CGTAGGCGTtaactccaggagTATATAAACTCAT
4	363_YDR281c/Phm6:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGGGTTTCTCGAGGAACTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGAG ACGTAGGCGTtaactccaggagTATATAAACTCAT
4	364_YLR268w/Sec22:TAvariants_CGTAGGCGTAGA_0_1_	ggttctggatcaggttcaCGTAGGCGTAGATTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGA TCTTCCTCtaactccaggagTATATAAACTCAT
4	365_YGL098w/Use1:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaAAGAGTAAATTGAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAAT TATTCCGTAGAtaactccaggagTATATAAACTCAT

4	366_YbR056W-A/Mnc1:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCTAAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	367_YLR268w/Sec22:TAvariants__5_0_CGTAGA	ggttctggatcaggttcaGCGCAAAAGATCAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCCGTAGAtaactccaggagTATATAAACTCAT
4	368_YPL192c/Prm3:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGAGAATAAGGGTTCATTTTACCAAGGTGCCATTTTCGGTTCGTTCCTTGGTGCTGCTGTAACTACTGTTTTAAGTAACCTAGCAGTTAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	369_YMR183c/Sso2:TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGCAAGAAAAAACAAAATAAGATGTTTGATCATCTGCTTTATTATCTTTGCTATTGTTGTTGTCGTTGTGGTTGTTCCATCCGTTGTGAGACGTAGGCGTtaactccaggagTATATAAACTCAT
4	370_YbR056W-A/Mnc1:TAvariants_5_5_	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTAACTGTGGCGATTGCTGTTCAGCAATTGCGAACGTTTTATGTTGTCTGTGTCTAATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	371_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATGTTCTGTGTTGTCTAGATTGTTCAATTAACTGTTGTAACGCGTGTTGCTGTGCATGTTGTTGTGGCTTAATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	372_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATTGCCTGGGCTGTCTATTAGTTAACGCGATTGCATGTTGTTGTAACTGTTGTGATTGTTCATGTTGTTGTATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	373_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATTGTCTAATTGCGGTTGGCTGTTGTTCATGTTGTCTGTTAGCAAACTGTAACTGTGATTGTTGCTGTTGTATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	374_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTAACTGTGATGGCCTATGTCTGTGTTCATGCTGTTGTTGTGTTAACGCGTGTGCAATTTTATGTATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	375_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATCTATGTTCAGCGAACTGTAACTGTGTTTGTTGTGCATTATGTTGTTGTCTGGGCTGTATTGATTGTTGCATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	376_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATTGTAACTGTGTTTGTTGTATTTTATCATGTTGCCTGGATAACTGTCTATGTGGCGCATGTGCGTGTTGTATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	377_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATCTATGTGATTGTGCATGTTGTTTATGTAACGGCGCGGTTTCATGCAACTGTTGTTGTTGTTGTCTGATTATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	378_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATAACTGTTGCGTTCTATGTTCATGTGCACTGTTATGTTGTTGTTGTGATATTTGTGCGGGCAACTGTTGTATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	379_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATGCGCTGGTTAACTGTTGTTGTTGTGATTGTTGTCTAGGCTGCTGTTGTTCATGTTTAATTAACTGTGCAATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	380_YbR056W-A/Mnc1:shuffled	ggttctggatcaggttcaCCACCAAGAAATGATGTTGGCTCATGTGATGCGTTATGTTGCCTATGTTGTTGTTGTCTGTGTATTTGTTGTAACAACTGTGCAATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	381_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTTATGTGTTGCGTGTTGTTCAAACAACGCACTGCTAGGCTGTTGTATTTGTGATATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	382_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTGTTTGTGGCGCGTGTTGTTCATTACTACTGATTAACGATTGTTGTGCATGTAACATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	383_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTGCGTGTCTAAACTGTTGTCTGGCAGGCGTTGATAACTCATGTTGTATTTGTTTAATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	384_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTGTTTGTAACGATTGTTGTGGCCTAGCATTACTGGCGATTTGTTGTTCATGTAACATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	385_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTCTGTGTGCGCTATGTTGTGATGGCGTTAACTCATTAGCATGTTGTAACTGTATTATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	386_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTGTTTGTGATCTGTGTTGTATTTTAAACGGCGCACTATCATGTTGTAACTGTGCGATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	387_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTGCATGTCTAGTTTGTTGTCTGAACGATGGCGCGTTAATTTGTTGTAACTGTTCAATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	388_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTATTTGTGATGCATGTTGTGGCAACCTGGTTTTATCAGCGTGTTGTCTATGTAACATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	389_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTTTATGTGCGGCATGTTGTAACGATGGCTCACTGATTGTTTGTTGTAACTGTCTAATTGATCTATGTTGTtaactccaggagTATATAAACTCAT
4	390_YbR056W-A/Mnc1:shuffled_fixed_C	ggttctggatcaggttcaCCACCAAGAAATGATTGTTGTTGTTGTTGTGATTGTCTGGGCTGTTGTAACTTAAACGTTGCATCACTATGTTGTATTTGTGCGATTGATCTATGTTGTtaactccaggagTATATAAACTCAT

Table S 5. Primers.

Primer	Sequence	Description
TMDpool-fw	CAGAAGTTGATTTCTGAAGAAGACCTCGGTTCTGGATCAGGTTCA	forward primer to amplify TMD oligo pools and extend the homology arms
TMDpool-rev	ACTCTTTATTCCTACATAAGTAAATGAGTTTATATACTCCTGGAG	reverse primer to amplify TMD oligo pools and extend the homology arms
TMDamp-fw	CTACACGACGCTCTTCCGATCTN$_{0-5}$ATGGACGAATTGTACAAGGA	mixture of four primers with 0, 2, 3 or 5 degenerate (N) positions
TMDamp-rev	GACGTGTGCTCTTCCGATCTN$_{0-5}$CGAATTCAAGCTTAGATCTGATA	mixture of four primers with 0, 2, 3 or 5 degenerate (N) positions
P5_Scriptseq	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	*phosphorothiolate bond
P7-ix_Scriptseq	CAAGCAGAAGACGGCATACGAGAT_ix_GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	_ix_ indicates a 6-nucleotide plate-specific index; *phosphorothiolate bond
mNG-TMD-fw	TTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAGATGCGCACTTAACTTCGCATCTG	forward primer to amplify kanMX-GPDpr-mNG-TMD cassette
mNG-TMD-rev	TATATACACATGTATATATATCGTATGCTGCAGCTTTAAATAATCGGTGTCACTACCTTCCTTCGTTCTTCCTTC	reverse primer to amplify kanMX-GPDpr-mNG-TMD cassette

Table S 6. Individual Oligonucleotides encoding TMD variants.

Oligo	Sequence
YDL241w_shuffled_1	ggttctggatcaggttcaAGGCCTTCTCATTCTCTCTATTTCACAATAGTGTTTGTCATGTTGGCCCTGCTAAAAATTACCtaactccaggagTATATAAACTCAT
YDL241w_shuffled_2	ggttctggatcaggttcaAGGCCTTCTCATTCTCTATTCGCCCTCACCATAGTCCTGATGTATTTGAAATTTGTGATTACAtaactccaggagTATATAAACTCAT
Ubc6_shuffled_1	ggttctggatcaggttcaCCTAATGATAGTTCTTATGGTGTTATTTTTTTTGCTTTGTTTATTTCAATGCTTATGGTTTTGGGCATCAAAtaactccaggagTATATAAACTCAT
Ubc6_shuffled_2	ggttctggatcaggttcaCCTAATGATAGTTCTTTTGGCATTTCACTTATGGTTGGTTTTTATTTGGCTGTTTTGTTTATTATGATCAAAtaactccaggagTATATAAACTCAT
Scs2_shuffled_1	ggttctggatcaggttcaAATGAATCATCCAGCATCTTCTTGTTCGGTCTCTTAATGATAGGACTTGTTTTGGCAATAGTTTGGTACAGAtaactccaggagTATATAAACTCAT
Scs2_shuffled_2	ggttctggatcaggttcaAATGAATCATCCAGCTTGGGTATCATAGTTCTCTTCCTTATGTTAATATTCTTGGTTGCATGGTACGGAAGAtaactccaggagTATATAAACTCAT
Gos1_shuffled_1	ggttctggatcaggttcaAGAAGGAAGAAAAACTTGACCTTTACCTTCGCCGTAACAGCGTGTCTGTTGCTTATAATATTTACGTTTTGGtaactccaggagTATATAAACTCAT
Gos1_shuffled_2	ggttctggatcaggttcaAGAAGGAAGAAAAACTTTACGTTTCTTTTTACCGTATGTATATTCGCCATACTGACCGCGACATTGTTGTGGtaactccaggagTATATAAACTCAT
Vam3_shuffled_1	ggttctggatcaggttcaAACAAATGCGGTAAGACCGCTATAATGATAGTCGTGTTAGTATTGATCATTCTTGTATGCGTGGTTCTAAGTtaactccaggagTATATAAACTCAT
Vam3_shuffled_2	ggttctggatcaggttcaAACAAATGCGGTAAGGTAATTATAGTTGTGACCGTAGCTGTCGTGTTGATCTTAATGTGCCTTATACTAAGTtaactccaggagTATATAAACTCAT
Use1_TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaAAGAGTAAATTGAGTTACTTGTTTTATATTACTGTTTTCATTTTTATGATTCTCGGACTGGTGTTTACATTTATCATAATTCAATTATTCAGACGTAGGCGTtaactccaggagTATATAAACTCAT
Sec22_TAvariants__5_0_AGACGTAGGCGT	ggttctggatcaggttcaGCGCAAAAGATCAACTTCGATCTCTTGATCAGTCAATATGCTCCTATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCAGACGTAGGCGTtaactccaggagTATATAAACTCAT
Sec20_variant1	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTCGTATTTTCAAGCATGATGAGCTA**taa**ctccaggagTATATAAACTCAT
Sec20_variant2	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTCGTATTTTCAAGTTACCTGTTAAACTCGGCCATGATGAGCTAtaactccaggagTATATAAACTCAT
Sec20_var3	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTCGTATTTTCAAGggcagtggttctggatccCATGATGAGCTAtaactccaggagTATATAAACTCAT
Sec20_var4	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTCGTATTTTCAAGggcagtggttctCATGATGAGCTAtaactccaggagTATATAAACTCAT
Sec20_var5	ggttctggatcaggttcaCAAGAGAAACGAGATGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGCGTCGTATTTTCAAGggcagtCATGATGAGCTAtaactccaggagTATATAAACTCAT
Sec20_26V_Ncharge	ggttctggatcaggttcaCAAtctAAACGAaagGTCTATTTATCACTTGGGTTCCTCCTATGCTGCGTCTCGTGGGTTCTATGGGTTGTCGTAGTTGTCGTAGTTGTCGTACGTtaactccaggagTATATAAACTCAT
Sbh1_26V_Ncharge	ggttctggatcaggttcaCTAAGAGTAaagagaTTAGTTGTGTTGTTTCTAGCGGTCGGTTTCATCTTTTCTGTTGTTGCATTACATGTTATTTCTGTTGTCGTAGTTGTCAAAtaactccaggagTATATAAACTCAT
Phm6_26V_Ncharge	ggttctggatcaggttcaGGGagaCTCaagaagCTCATCATCGTCATTATTATTGTGTTGCTGTTGTACAGTTTGACAATGGTAGGTCTTTTTTATGTGGTTGTCGTAGTTATGtaactccaggagTATATAAACTCAT
Scs2_26V_Ncharge	ggttctggatcaggttcaAATagaTCAaagaagATGGGTATATTCATATTGGTTGCACTCCTTATCTTGGTTTTAGGATGGTTCTACGTTGTCGTAGTTGTCGTAGTTGTCAGAtaactccaggagTATATAAACTCAT
Bos1_26V_Ncharge	ggttctggatcaggttcaGTGTTCAAAagaAAACTAGTCTTTTGGATCGCGTTAATTCTCTTGATCATAGGTATTTATTATGTGTTGGTTGTCGTAGTTGTCGTAGTTGTCAAAtaactccaggagTATATAAACTCAT

Table S 7. List of general consumables, machines and software used.

General solutions or chemicals

Item	Supplier
Velocity DNA polymerase	Meridian Bioscience
NEBNextUltra DNA library prep kit	New England Biolabs
cytosine	BIC0713, Apollo Scientific
5-fluorocytosine	PC3735, Apollo Scientific
Monosodium glutamic acid (MSG)	G1626, Sigma-Aldrich
Canavanine	C1625, Sigma-Aldrich
Thialysine	A2636, Sigma-Aldrich
G418	A291-25, Biochrom
Hygromycin B Gold	ant-hg-5, Invivogen
clonNAT	5.0, Werner BioAgents
Adenine	A8626, Sigma-Aldrich
Concanavalin A	C2010, Sigma-Aldrich

Bacto yeast extract	Thermo Fisher Scientific
Bacto peptone	Thermo Fisher Scientific
Bacto agar	Thermo Fisher Scientific
Bacto yeast nitrogen base without amino acids	BD Biosciences
Bacto yeast nitrogen base without amino acids and ammonium sulfate	BD Biosciences
Yeast Nitrogen base without Amino acids and without Folic Acid and Riboflavin. LoFlo	FORMEDIUM
Gentra Puregene Yeast/Bact. Kit B	Qiagen
Plasmid Miniprep Kit	Qiagen
QIAquick gel extraction kit	Qiagen
MiSeq Standard FC (300 cycles) v2	Illumina
NextSeq 500/550 Mid Output Kit v2.5 (300 cycles)	Illumina
Ethanol	Sigma Aldrich
Sodium chloride (NaCl)	Sigma Aldrich
PhenoPlate 384-well microplates	PerkinElmer
PhenoPlate 96-well microplates	PerkinElmer

Table S 8. List of software used.

Software		
Item	**Developer**	**Identifier**
Adobe Illustrator 24.3	Adobe	
Microsoft Excel	Microsoft	
R studio 1.2.5033	R studio PBC	
ImageJ	(Schneider et al., 2012)	https://imagej.nih.gov/ij/
ImageLab	Bio-Rad	www.bio-rad.com/en-hk/product/image-lab-software?II
Phobius	(Käll et al., 2004)	
SignalP	(Mitchell et al., 2019)	
NLStradamus	(Nguyen Ba et al., 2009)	https://www.moseslab.csb.utoronto.ca/NLStradamus/
TMHMM	(Krogh et al., 2001)	

Table S 9. List of machines used.

Machines	
Item	**Supplier**
ROTOR	Singer Instruments
PIXL	Singer Instruments
PhenoBooth	Singer Instruments
Opera Phenix	PerkinElmer
Infinite M1000 plate reader	Tecan
ChemiDoc imaging system	BioRad
Thermoshaker	Eppendorf
Professional Trio Thermocycler	Biometra

8.2 LIST OF ABBREVIATIONS

AA – amino acid

AAA - ATPase associated with diverse cellular activities

ADE – adenine

CRISPR - clustered regularly-interspaced short palindromic repeats

CFTR - cystic fibrosis transmembrane conductance regulator

CYSTM - cysteine-rich transmembrane module

DMS - deep mutational scanning

DNA - deoxyribonucleic acid

DSB – double strain breaks

DUB – deubiquitinase

EMC - ER-membrane protein complex

ER - endoplasmic reticulum

ERAD - ER-associated degradation

ER-SURF - ER surface retrieval pathway

FC – fluorocytosine

FHH - familial hypocalciuric hypercalcemia

FISH - fluorescent oligonucleotide probes

FU – fluorouracil

Gal – galactose

GET - guided entry of tail-anchored protein

GPI - glycosylphosphatidylinositol

HR – homologous recombination

IACS - image-activated cell sorting

iPAL - imaging pooled-to-arrayed libraries

MAD - mitochondria-associated degradation

MPP - mitochondrial processing peptidase

MRP - mitochondrial ribosomal protein

MSF - mitochondrial import stimulation factor

MTS - mitochondria-targeting sequence

mNG – mNeonGreen

mSc-I – mScarletI

NE - nuclear envelope

NGS – next-generation sequencing

NHEJ - non-homologous end joining

NLS - nuclear localization signal

NoS - nucleolar localization signal

NSHPT - neonatal severe hyperparathyroidism

OMM - the outer mitochondrial membrane

ORF – open reading frame

PI4P - phosphatidylinositol 4-phosphate

PM – plasma membrane

PTS - peroxisomal targeting signal

Raf – raffinose

QC – quality control

RNA – ribonucleic acid

RNC - ribosome-nascent chain complex

SR - SRP receptor

SRP - signal recognition particle

SPC - signal peptidase complex

TA - tail-anchored

TIM - translocase of the inner membrane

TMD - transmembrane domain

TOM - translocase of the outer membrane

TPR – tetratricopeptide

v-SNARE - vesicle membrane receptor protein

UBQLNs - ubiquilin family proteins

UPS – ubiquitin-proteasome system

URA – uracil

9 REFERENCES

9.1 REFERENCES IN ALPHABETICAL ORDER

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). TensorFlow: a system for large-scale machine learning. Proc. 12th USENIX Conf. Oper. Syst. Des. Implement. 265–283. https://doi.org/https://doi.org/10.48550/arXiv.1605.08695.

Ali, M.A.S., Misko, O., Salumaa, S.O., Papkov, M., Palo, K., Fishman, D., and Parts, L. (2021). Evaluating Very Deep Convolutional Neural Networks for Nucleus Segmentation from Brightfield Cell Microscopy Images. SLAS Discov. *26*, 1125–1137. https://doi.org/10.1177/24725552211023214.

Andreeva, N., Kulakovskaya, E., Zvonarev, A., Penin, A., Eliseeva, I., Teterina, A., Lando, A., Kulakovskiy, I. V., and Kulakovskaya, T. (2017). Transcriptome profile of yeast reveals the essential role of PMA2 and uncharacterized gene YBR056W-A (MNC1) in adaptation to toxic manganese concentration. Metallomics *9*, 175–182. https://doi.org/10.1039/c6mt00210b.

Antonsson, B. (2001). Bax and other pro-apoptotic Bcl-2 family "killer-proteins" and their victim the mitochondrion. Cell Tissue Res. *306*, 347–361. https://doi.org/10.1007/s00441-001-0472-0.

Arndt, V., Rogon, C., and Höhfeld, J. (2007). To be, or not to be — molecular chaperones in protein degradation. Cell. Mol. Life Sci. *64*, 2525–2541. https://doi.org/10.1007/S00018-007-7188-6.

Arnold, A., Horst, S.A., Gardella, T.J., Baba, H., Levine, M.A., and Kronenberg, H.M. (1990). Mutation of the signal peptide-encoding region of the preproparathyroid hormone gene in familial isolated hypoparathyroidism. J. Clin. Invest. *86*, 1084–1087. https://doi.org/10.1172/JCI114811.

Ast, T., Cohen, G., and Schuldiner, M. (2013). A Network of Cytosolic Factors Targets SRP-Independent Proteins to the Endoplasmic Reticulum. Cell *152*, 1134–1145. https://doi.org/10.1016/J.CELL.2013.02.003.

Aviram, N., Ast, T., Costa, E.A., Arakel, E.C., Chuartzman, S.G., Jan, C.H., Haßdenteufel, S., Dudek, J., Jung, M., Schorr, S., et al. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. Nature *540*, 134–138. https://doi.org/10.1038/nature20169.

Bai, L., You, Q., Feng, X., Kovach, A., and Li, H. (2020). Structure of the ER membrane complex, a transmembrane-domain insertase. Nature *584*, 475–478. https://doi.org/10.1038/s41586-020-2389-3.

Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F.J., Myers, C.L., Andrews, B., and Boone, C. (2010). Synthetic Genetic Array (SGA) Analysis in Saccharomyces cerevisiae and Schizosaccharomyces pombe. Methods Enzymol. *470*, 145–179. https://doi.org/10.1016/S0076-6879(10)70007-0.

Bays, N.W., Gardner, R.G., Seelig, L.P., Joazeiro, C.A., and Hampton, R.Y. (2000). Hrd1p/Der3p is a membrane-anchored ubiquitin ligase required for ER-associated

degradation. Nat. Cell Biol. *3*, 24–29. https://doi.org/10.1038/35050524.

Beilharz, T., Egan, B., Silver, P.A., Hofmann, K., and Lithgow, T. (2003). Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in Saccharomyces cerevisiae. J. Biol. Chem. *278*, 8219–8223. https://doi.org/10.1074/jbc.M212725200.

Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., et al. (2019). ilastik: interactive machine learning for (bio)image analysis. Nat. Methods *16*, 1226–1232. https://doi.org/10.1038/s41592-019-0582-9.

Bonifacino, J.S., and Glick, B.S. (2004). The Mechanisms of Vesicle Budding and Fusion. Cell *116*, 153–166. https://doi.org/10.1016/S0092-8674(03)01079-1.

Borgese, N., Coy-Vergara, J., Colombo, S.F., and Schwappach, B. (2019). The Ways of Tails: the GET Pathway and more. Protein J. *38*, 289–305. https://doi.org/10.1007/s10930-019-09845-4.

Boutros, M., Heigwer, F., and Laufer, C. (2015). Microscopy-Based High-Content Screening. Cell *163*, 1314–1325. https://doi.org/10.1016/j.cell.2015.11.007.

Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and 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*, 115–132. https://doi.org/https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2%3C115::AID-YEA204%3E3.0.CO;2-2.

Brameier, M., Krings, A., and MacCallum, R.M. (2007). NucPred—Predicting nuclear localization of proteins. Bioinformatics *23*, 1159–1160. https://doi.org/10.1093/BIOINFORMATICS/BTM066.

Brocard, C., and Hartig, A. (2006). Peroxisome targeting signal 1: Is it really a simple tripeptide? Biochim. Biophys. Acta - Mol. Cell Res. *1763*, 1565–1573. https://doi.org/10.1016/j.bbamcr.2006.08.022.

Brown, D.A., and London, E. (1998). Structure and origin of ordered lipid domains in biological membranes. J. Membr. Biol. *164*, 103–114. https://doi.org/10.1007/S002329900397.

Buchberger, A., Bukau, B., and Sommer, T. (2010). Protein Quality Control in the Cytosol and the Endoplasmic Reticulum: Brothers in Arms. Mol. Cell *40*, 238–252. https://doi.org/10.1016/J.MOLCEL.2010.10.001.

Burri, L., and Lithgow, T. (2004). A Complete Set of SNAREs in Yeast. Traffic *5*, 45–52. https://doi.org/10.1046/j.1600-0854.2003.00151.x.

Bykov, Y.S., Flohr, T., Boos, F., Zung, N., Herrmann, J.M., and Schuldiner, M. (2022). Widespread use of unconventional targeting signals in mitochondrial ribosome proteins. EMBO J. *41*, e109519. https://doi.org/10.15252/EMBJ.2021109519.

Carvalho, P., Goder, V., and Rapoport, T.A. (2006). Distinct Ubiquitin-Ligase Complexes Define Convergent Pathways for the Degradation of ER Proteins. Cell *126*, 361–373. https://doi.org/10.1016/J.CELL.2006.05.043.

Cassanelli, S., Bertolini, S., Rolleri, M., De Stefano, F., Casarino, L., Elicio, N., Naselli, A., and Calandra, S. (1998). A “de novo” point mutation of the low-density lipoprotein receptor gene in an Italian subject with primary hypercholesterolemia. Clin. Genet. *53*, 391–395. https://doi.org/10.1111/j.1399-0004.1998.tb02752.x.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*. https://doi.org/10.1126/science.aaa6090.

Chu, C.T., Plowey, E.D., Wang, Y., Patel, V., and Jordan-Sciutto, K.L. (2007). Location, Location, Location: Altered Transcription Factor Trafficking in Neurodegeneration. J. Neuropathol. Exp. Neurol. *66*, 873–883. https://doi.org/10.1097/NEN.0B013E318156A3D7.

Costa, E.A., Subramanian, K., Nunnari, J., and Weissman, J.S. (2018). Defining the physiological role of SRP in protein-targeting efficiency and specificity. Science *359*, 689–692. https://doi.org/10.1126/science.aar3607.

Da Costa, L., Tchernia, G., Gascard, P., Lo, A., Meerpohl, J., Niemeyer, C., Chasis, J.A., Fixler, J., and Mohandas, N. (2003). Nucleolar localization of RPS19 protein in normal cells and mislocalization due to mutations in the nucleolar localization signals in 2 Diamond-Blackfan anemia patients: potential insights into pathophysiology. Blood *101*, 5039–5045. https://doi.org/10.1182/BLOOD-2002-12-3878.

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., et al. (2017). Predicting the targeting of tail-anchored proteins to subcellular compartments in mammalian cells. J. Cell Sci. *130*, 1675–1687. https://doi.org/10.1242/jcs.200204.

Culver, J.A., and Mariappan, M. (2021). Deubiquitinases USP20/33 promote the biogenesis of tail-anchored membrane proteins. J. Cell Biol. *220*, e202004086. https://doi.org/10.1083/jcb.202004086.

Culver, J.A., Li, X., Jordan, M., and Mariappan, M. (2022). A second chance for protein targeting/folding: Ubiquitination and deubiquitination of nascent proteins. BioEssays *44*, 2200014. https://doi.org/10.1002/BIES.202200014.

Curran, K.A., Morse, N.J., Markham, K.A., Wagman, A.M., Gupta, A., and Alper, H.S. (2015). Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. ACS Synth. Biol. *4*, 824–832. https://doi.org/10.1021/sb5003357.

Deak, P.M., and Wolf, D.H. (2001). Membrane Topology and Function of Der3/Hrd1p as a Ubiquitin-Protein Ligase (E3) Involved in Endoplasmic Reticulum Degradation. J. Biol. Chem. *276*, 10663–10669. https://doi.org/10.1074/JBC.M008608200.

Dederer, V., and Lemberg, M.K. (2021). Transmembrane dislocases: a second chance for protein targeting. Trends Cell Biol. *31*, 898–911. https://doi.org/10.1016/J.TCB.2021.05.007.

Dederer, V., Khmelinskii, A., Huhn, A.G., Okreglak, V., Knop, M., and Lemberg, M.K. (2019). Cooperation of mitochondrial and ER factors in quality control of tail-anchored proteins. Elife *8*, e45506. https://doi.org/10.7554/eLife.45506.

Denic, V. (2012). A portrait of the GET pathway as a surprisingly complicated young man.

Trends Biochem. Sci. *37*, 411–417. https://doi.org/10.1016/j.tibs.2012.07.004.

Djordjevic, S., Zhang, X., Bartlam, M., Ye, S., Rao, Z., and Danpure, C.J. (2010). Structural implications of a G170R mutation of alanine:glyoxylate aminotransferase that is associated with peroxisome-to-mitochondrion mistargeting. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. *66*, 233. https://doi.org/10.1107/S1744309109054645.

Ear, P.H., and Michnick, S.W. (2009). A general life-death selection strategy for dissecting protein functions. Nat. Methods *6*, 813–816. https://doi.org/10.1038/nmeth.1389.

Emanuel, G., Moffitt, J.R., and Zhuang, X. (2017). High-throughput, image-based screening of pooled genetic-variant libraries. Nat. Methods *14*, 1159–1162. https://doi.org/10.1038/nmeth.4495.

Eng, C.H.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.C., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature *568*, 235–239. https://doi.org/10.1038/s41586-019-1049-y.

Fairn, G.D., and Grinstein, S. (2012). Precursor or charge supplier? Science *337*, 653–654. https://doi.org/10.1126/SCIENCE.1227096.

Feldman, D., Singh, A., Schmid-Burgk, J.L., Carlson, R.J., Mezger, A., Garrity, A.J., Zhang, F., and Blainey, P.C. (2019). Optical Pooled Screens in Human Cells. Cell *179*, 787-799.e17. https://doi.org/10.1016/j.cell.2019.09.016.

Figueiredo Costa, B., Cassella, P., Colombo, S.F., and Borgese, N. (2018). Discrimination between the endoplasmic reticulum and mitochondria by spontaneously inserting tail-anchored proteins. Traffic *19*, 182–197. https://doi.org/10.1111/tra.12550.

Foresti, O., Rodriguez-Vaello, V., Funaya, C., and Carvalho, P. (2014). Quality control of inner nuclear membrane proteins by the Asi complex. Science *346*, 751–755. https://doi.org/10.1126/SCIENCE.1255638.

Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807. https://doi.org/10.1038/nmeth.3027.

Fry, M.Y., Saladi, S.M., Cunha, A., and Clemons, W.M. (2021). Sequence-based features that are determinant for tail-anchored membrane protein sorting in eukaryotes. Traffic *22*, 306–318. https://doi.org/10.1111/TRA.12809.

Gilmore, R., Blobel, G., and Walter, P. (1982). Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. J. Cell Biol. *95*, 463. https://doi.org/10.1083/JCB.95.2.463.

Goder, V., Crottet, P., and Spiess, M. (2000). In vivo kinetics of protein targeting to the endoplasmic reticulum determined by site-specific phosphorylation. EMBO J. *19*, 6704. https://doi.org/10.1093/EMBOJ/19.24.6704.

Guna, A., and Hegde, R.S. (2018). Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. Curr. Biol. *28*, R498–R511. https://doi.org/10.1016/j.cub.2018.02.004.

Guna, A., Volkmar, N., Christianson, J.C., and Hegde, R.S. (2018). The ER membrane protein

complex is a transmembrane domain insertase. Science *359*, 470–473. https://doi.org/10.1126/science.aao3099.

Güngör, B., Flohr, T., Garg, S.G., and Herrmann, J.M. (2022). The ER membrane complex (EMC) can functionally replace the Oxa1 insertase in mitochondria. PLOS Biol. *20*, e3001380. https://doi.org/10.1371/JOURNAL.PBIO.3001380.

Guo, H., Xiong, Y., Witkowski, P., Cui, J., Wang, L.J., Sun, J., Lara-Lemus, R., Haataja, L., Hutchison, K., Shan, S.O., et al. (2014). Inefficient translocation of preproinsulin contributes to pancreatic β cell failure and late-onset diabetes. J. Biol. Chem. *289*, 16290–16302. https://doi.org/10.1074/jbc.M114.562355.

Habib, S.J., Vasiljev, A., Neupert, W., and Rapaport, D. (2003). Multiple functions of tail-anchor domains of mitochondrial outer membrane proteins. FEBS Lett. *555*, 511–515. https://doi.org/10.1016/S0014-5793(03)01325-5.

Habib, S.J., Neupert, W., and Rapaport, D. (2007). Analysis and Prediction of Mitochondrial Targeting Signals. Methods Cell Biol. *80*, 761–781. https://doi.org/10.1016/S0091-679X(06)80035-X.

Hammond, G.R.V., Fischer, M.J., Anderson, K.E., Holdich, J., Koteci, A., Balla, T., and Irvine, R.F. (2012). PI4P and PI(4,5)P2 are essential but independent lipid determinants of membrane identity. Science *337*, 727–730. https://doi.org/10.1126/science.1222483.

Hanna, R.E., and Doench, J.G. (2020). Design and analysis of CRISPR–Cas experiments. Nat. Biotechnol. *38*, 813–823. https://doi.org/10.1038/s41587-020-0490-7.

Hansen, K.G., Aviram, N., Laborenz, J., Bibi, C., Meyer, M., Spang, A., Schuldiner, M., and Herrmann, J.M. (2018). An ER surface retrieval pathway safeguards the import of mitochondrial membrane proteins in yeast. Science *361*, 1118–1122. https://doi.org/10.1126/science.aar8174.

Hartzog, P.E., Nicholson, B.P., and McCusker, J.H. (2005). Cytosine deaminase MX cassettes as positive/negative selectable markers in Saccharomyces cerevisiae. Yeast *22*, 789–798. https://doi.org/10.1002/yea.1245.

Hasle, N., Cooke, A., Srivatsan, S., Huang, H., Stephany, J.J., Krieger, Z., Jackson, D., Tang, W., Pendyala, S., Monnat, R.J., et al. (2020). High-throughput, microscope-based sorting to dissect cellular heterogeneity. Mol. Syst. Biol. *16*. https://doi.org/10.15252/msb.20209442.

Haßdenteufel, S., Sicking, M., Schorr, S., Aviram, N., Fecher-Trost, C., Schuldiner, M., Jung, M., Zimmermann, R., and Lang, S. (2017). hSnd2 protein represents an alternative targeting factor to the endoplasmic reticulum in human cells. FEBS Lett. *591*, 3211–3224. https://doi.org/10.1002/1873-3468.12831.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. *2016*, 770–778. https://doi.org/10.1109/CVPR.2016.90.

Hegde, R.S., and Keenan, R.J. (2011). Tail-anchored membrane protein insertion into the endoplasmic reticulum. Nat. Rev. Mol. Cell Biol. *12*, 787–798. https://doi.org/10.1038/nrm3226.

Hegde, R.S., and Keenan, R.J. (2021). The mechanisms of integral membrane protein biogenesis. Nat. Rev. Mol. Cell Biol. *23*, 107–124. https://doi.org/10.1038/s41580-021-00413-2.

Hegde, R.S., and Zavodszky, E. (2019). Recognition and Degradation of Mislocalized Proteins in Health and Disease. Cold Spring Harb. Perspect. Biol. *11*, a033902. https://doi.org/10.1101/cshperspect.a033902.

von Heijne, G., and Abrahmsèn, L. (1989). Species-specific variation in signal peptide design Implications for protein secretion in foreign hosts. FEBS Lett. *244*, 439–446. https://doi.org/10.1016/0014-5793(89)80579-4.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature *433*, 377–381. https://doi.org/10.1038/nature03216.

Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., and von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature *450*, 1026–1030. https://doi.org/10.1038/nature06387.

Holthuis, J.C.M., Pomorski, T., Raggers, R.J., Sprong, H., and Van Meer, G. (2001). The organizing potential of sphingolipids in intracellular membrane transport. Physiol. Rev. *81*, 1689–1723. https://doi.org/10.1152/PHYSREV.2001.81.4.1689.

De Hoop, M.J., and Ab, G. (1992). Import of proteins into peroxisomes and other microbodies. Biochem. J. *286*, 657. https://doi.org/10.1042/BJ2860657.

Hung, M.C., and Link, W. (2011). Protein localization in disease and therapy. J. Cell Sci. *124*, 3381–3392. https://doi.org/10.1242/jcs.089110.

Hussain, S., Mohd Ali, J., Jalaludin, M.Y., and Harun, F. (2013). Permanent neonatal diabetes due to a novel insulin signal peptide mutation. Pediatr. Diabetes *14*, 299–303. https://doi.org/10.1111/pedi.12011.

Ingolia, N.T., Hussmann, J.A., and Weissman, J.S. (2019). Ribosome Profiling: Global Views of Translation. Cold Spring Harb. Perspect. Biol. *11*, a032698. https://doi.org/10.1101/CSHPERSPECT.A032698.

Itakura, E., Zavodszky, E., Shao, S., Wohlever, M.L., Keenan, R.J., and Hegde, R.S. (2016). Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation. Mol. Cell *63*, 21–33. https://doi.org/10.1016/J.MOLCEL.2016.05.020.

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., et al. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast *21*, 947–962. https://doi.org/10.1002/yea.1142.

Jarjanazi, H., Savas, S., Pabalan, N., Dennis, J.W., and Ozcelik, H. (2007). Biological implications of SNPs in signal peptide domains of human proteins. Proteins Struct. Funct. Bioinforma. *70*, 394–403. https://doi.org/10.1002/prot.21548.

Jarjanazi, H., Savas, S., Pabalan, N., Dennis, J.W., and Ozcelik, H. (2008). Biological

implications of SNPs in signal peptide domains of human proteins. Proteins *70*, 394–403. https://doi.org/10.1002/PROT.21548.

Jomaa, A., Gamerdinger, M., Hsieh, H.-H., Wallisch, A., Chandrasekaran, V., Ulusoy, Z., Scaiola, A., Hegde, R.S., Shan, S., Ban, N., et al. (2022). Mechanism of signal sequence handover from NAC to SRP on ribosomes during ER-protein targeting. Science *375*, 839–844. https://doi.org/10.1126/SCIENCE.ABL6459.

Joshi, J.R., Singh, V., and Friedman, H. (2020). Arabidopsis cysteine-rich trans-membrane module (CYSTM) small proteins play a protective role mainly against heat and UV stresses. Funct. Plant Biol. *47*, 195. https://doi.org/10.1071/FP19236.

Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A short amino acid sequence able to specify nuclear location. Cell *39*, 499–509. https://doi.org/10.1016/0092-8674(84)90457-4.

Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. *338*, 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016.

Kanfer, G., Sarraf, S.A., Maman, Y., Baldwin, H., Dominguez-Martin, E., Johnson, K.R., Ward, M.E., Kampmann, M., Lippincott-Schwartz, J., and Youle, R.J. (2021). Image-based pooled whole-genome CRISPRi screening for subcellular phenotypes. J. Cell Biol. *220*. https://doi.org/10.1083/JCB.202006180.

Karaplis, A.C., Lim, S.K., Baba, H., Arnold, A., and Kronenberg, H.M. (1995). Inefficient membrane targeting, translocation, and proteolytic processing by signal peptidase of a mutant preproparathyroid hormone protein. J. Biol. Chem. *270*, 1629–1635. https://doi.org/10.1074/jbc.270.4.1629.

Kaufmann, T., Schlipf, S., Sanz, J., Neubert, K., Stein, R., and Borner, C. (2003). Characterization of the signal that directs Bcl-x(L), but not Bcl-2, to the mitochondrial outer membrane. J. Cell Biol. *160*, 53–64. https://doi.org/10.1083/jcb.200210084.

Keskin, A., Akdoğan, E., and Dunn, C.D. (2017). Evidence for Amino Acid Snorkeling from a High-Resolution, In Vivo Analysis of Fis1 Tail-Anchor Insertion at the Mitochondrial Outer Membrane. Genetics *205*, 691–705. https://doi.org/10.1534/genetics.116.196428.

Khmelinskii, A., and Knop, M. (2014). Analysis of Protein Dynamics with Tandem Fluorescent Protein Timers. Methods Mol. Biol. *1174*, 195–210. https://doi.org/10.1007/978-1-4939-0944-5_13.

Khmelinskii, A., Blaszczak, E., Pantazopoulou, M., Fischer, B., Omnus, D.J., Le Dez, G., Brossard, A., Gunnarsson, A., Barry, J.D., Meurer, M., et al. (2014). Protein quality control at the inner nuclear membrane. Nature *516*, 410–413. https://doi.org/10.1038/nature14096.

Kingma, D.P., and Ba, J.L. (2015). Adam: A method for stochastic optimization. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. https://doi.org/10.48550/arxiv.1412.6980.

Kong, K.-Y.E., Fischer, B., Meurer, M., Kats, I., Li, Z., Rühle, F., Barry, J.D., Kirrmaier, D., Chevyreva, V., San Luis, B.-J., et al. (2021a). Timer-based proteomic profiling of the ubiquitin-proteasome system reveals a substrate receptor of the GID ubiquitin ligase. Mol. Cell *81*.

https://doi.org/10.1016/j.molcel.2021.04.018.

Kong, K.Y.E., Coelho, J.P.L., Feige, M.J., and Khmelinskii, A. (2021b). Quality control of mislocalized and orphan proteins. Exp. Cell Res. *403*, 112617. https://doi.org/10.1016/J.YEXCR.2021.112617.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580. https://doi.org/10.1006/JMBI.2000.4315.

Krumpe, K., Frumkin, I., Herzig, Y., Rimon, N., Özbalci, C., Brügger, B., Rapaport, D., and Schuldiner, M. (2012). Ergosterol content specifies targeting of tail-anchored proteins to mitochondrial outer membranes. Mol. Biol. Cell *23*, 3927–3935. https://doi.org/10.1091/mbc.E11-12-0994.

Kumamoto, T., Morohashi, K. ichirou, Ito, A., and Omura, T. (1987). Site-directed mutagenesis of basic amino acid residues in the extension peptide of P-450(SCC) precursor: effects on the import of the precursor into mitochondria. J. Biochem. *102*, 833–838. https://doi.org/10.1093/OXFORDJOURNALS.JBCHEM.A122122.

Lanza, F., De La Salle, C., Baas, M.J., Schwartz, A., Boval, B., Cazenave, J.P., and Caen, J.P. (2002). A Leu7Pro mutation in the signal peptide of platelet glycoprotein (GP)IX in a case of Bernard-Soulier syndrome abolishes surface expression of the GPIb-V-IX complex. Br. J. Haematol. *118*, 260–266. https://doi.org/10.1046/J.1365-2141.2002.03544.X.

Lashuel, H.A., Hartley, D., Petre, B.M., Walz, T., and Lansbury, P.T. (2002). Amyloid pores from pathogenic mutations. Nature *418*, 291–291. https://doi.org/10.1038/418291a.

Lazarow, P.B. (2006). The import receptor Pex7p and the PTS2 targeting sequence. Biochim. Biophys. Acta *1763*, 1599–1604. https://doi.org/10.1016/J.BBAMCR.2006.08.011.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. Science *343*, 1360–1363. https://doi.org/10.1126/science.1250212.

Lewis, B.A., and Engelman, D.M. (1983). Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. J. Mol. Biol. *166*, 211–217. https://doi.org/10.1016/S0022-2836(83)80007-2.

Lewis, M.J., Nichols, B.J., Prescianotto-Baschong, C., Riezman, H., and Pelham, H.R.B. (2000). Specific retrieval of the exocytic SNARE Snc1p from early yeast endosomes. Mol. Biol. Cell *11*, 23–38. https://doi.org/10.1091/mbc.11.1.23.

Li, L., Zheng, J., Wu, X., and Jiang, H. (2019). Mitochondrial AAA-ATPase Msp1 detects mislocalized tail-anchored proteins through a dual-recognition mechanism. EMBO Rep. *20*, e46989. https://doi.org/10.15252/embr.201846989.

Lin, J. rong, and Hu, J. (2013). SeqNLS: Nuclear Localization Signal Prediction Based on Frequent Pattern Mining and Linear Motif Scoring. PLoS One *8*, e76864. https://doi.org/10.1371/JOURNAL.PONE.0076864.

Lin, K.F., Fry, M.Y., Saladi, S.M., and Clemons, W.M. (2021). Molecular basis of tail-anchored

integral membrane protein recognition by the cochaperone Sgt2. J. Biol. Chem. *296*. https://doi.org/10.1016/J.JBC.2021.100441.

Liu, J.J. (2016). Retromer-Mediated Protein Sorting and Vesicular Trafficking. J. Genet. Genomics *43*, 165–177. https://doi.org/10.1016/J.JGG.2016.02.006.

El Magraoui, F., Brinkmeier, R., Mastalski, T., Hupperich, A., Strehl, C., Schwerter, D., Girzalsky, W., Meyer, H.E., Warscheid, B., Erdmann, R., et al. (2019). The deubiquitination of the PTS1-import receptor Pex5p is required for peroxisomal matrix protein import. Biochim. Biophys. Acta - Mol. Cell Res. *1866*, 199–213. https://doi.org/10.1016/j.bbamcr.2018.11.002.

Manford, A.G., Stefan, C.J., Yuan, H.L., MacGurn, J.A., and Emr, S.D. (2012). ER-to-Plasma Membrane Tethering Proteins Regulate Cell Signaling and ER Morphology. Dev. Cell *23*, 1129–1140. https://doi.org/10.1016/J.DEVCEL.2012.11.004.

Marshansky, V., and Futai, M. (2008). The V-type H+-ATPase in vesicular trafficking: targeting, regulation and function. Curr. Opin. Cell Biol. *20*, 415. https://doi.org/10.1016/J.CEB.2008.03.015.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10. https://doi.org/10.14806/ej.17.1.200.

Mateja, A., and Keenan, R.J. (2018). A structural perspective on tail-anchored protein biogenesis by the GET pathway. Curr. Opin. Struct. Biol. *51*, 195–202. https://doi.org/10.1016/J.SBI.2018.07.009.

Matsumoto, S., Nakatsukasa, K., Kakuta, C., Tamura, Y., Esaki, M., and Endo, T. (2019). Msp1 Clears Mistargeted Proteins by Facilitating Their Transfer from Mitochondria to the ER. Mol. Cell *76*, 191–205. https://doi.org/10.1016/j.molcel.2019.07.006.

Matsumoto, S., Ono, S., Shinoda, S., Kakuta, C., Okada, S., Ito, T., Numata, T., and Endo, T. (2022). GET pathway mediates transfer of mislocalized tail-anchored proteins from mitochondria to the ER. J. Cell Biol. *221*, e202104076. https://doi.org/10.1083/JCB.202104076.

Mattiazzi Usaj, M., Styles, E.B., Verster, A.J., Friesen, H., Boone, C., and Andrews, B.J. (2016). High-Content Screening for Quantitative Cell Biology. Trends Cell Biol. *26*, 598–611. https://doi.org/10.1016/j.tcb.2016.03.008.

Mckenna, M.J., Sim, S.I., Ordureau, A., Wei, L., Harper, J.W., Shao, S., and Park, E. (2020). The endoplasmic reticulum P5A-ATPase is a transmembrane helix dislocase. Science *369*. https://doi.org/10.1126/science.abc5809.

McLane, L.M., and Corbett, A.H. (2009). Nuclear localization signals and human disease. IUBMB Life *61*, 697–706. https://doi.org/10.1002/IUB.194.

McQuown, A.J., Reif, D., and Denic, V. (2021). A TRCky TA protein delivery service snubs the UPS. J. Cell Biol. *220*, e202103196. https://doi.org/10.1083/jcb.202103196.

Meurer, M., Duan, Y., Sass, E., Kats, I., Herbst, K., Buchmuller, B.C., Dederer, V., Huber, F., Kirrmaier, D., Štefl, M., et al. (2018). Genome-wide C-SWAT library for high-throughput yeast genome tagging. Nat. Methods *15*, 598–600. https://doi.org/10.1038/s41592-018-0045-8.

Milanesi, L., Sheynis, T., Xue, W.F., Orlova, E. V., Hellewell, A.L., Jelinek, R., Hewitt, E.W., Radford, S.E., and Saibil, H.R. (2012). Direct three-dimensional visualization of membrane disruption by amyloid fibrils. PNAS *109*, 20455–20460. https://doi.org/10.1073/PNAS.1206325109.

Miller-Vedam, L.E., Bräuning, B., Popova, K.D., Schirle Oakdale, N.T., Bonnar, J.L., Prabu, J.R., Boydston, E.A., Sevillano, N., Shurtleff, M.J., Stroud, R.M., et al. (2020). Structural and mechanistic basis of the EMC-dependent biogenesis of distinct transmembrane clients. Elife *9*, e62611. https://doi.org/10.7554/eLife.62611.

Mir, R., and León, J. (2014). Pathogen and Circadian Controlled 1 (PCC1) Protein Is Anchored to the Plasma Membrane and Interacts with Subunit 5 of COP9 Signalosome in Arabidopsis. PLoS One *9*, e87216. https://doi.org/10.1371/JOURNAL.PONE.0087216.

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., et al. (2019). InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. *47*, D351–D360. https://doi.org/10.1093/nar/gky1100.

Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G., and Engelman, D.M. (2004). Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. PNAS *101*, 4083–4088. https://doi.org/10.1073/PNAS.0307332101.

Mizutani, A., Matsuzaki, A., Momoi, M.Y., Fujita, E., Tanabe, Y., and Momoi, T. (2007). Intracellular distribution of a speech/language disorder associated FOXP2 mutant. Biochem. Biophys. Res. Commun. *353*, 869–874. https://doi.org/10.1016/J.BBRC.2006.12.130.

Mumberg, D., Müller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene *156*, 119–122. https://doi.org/https://doi.org/10.1016/0378-1119(95)00037-7.

Natarajan, N., Foresti, O., Wendrich, K., Stein, A., and Carvalho, P. (2020). Quality Control of Protein Complex Assembly by a Transmembrane Recognition Factor. Mol. Cell *77*, 108–119. https://doi.org/10.1016/j.molcel.2019.10.003.

Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A., and Eisenhaber, F. (2003). Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J. Mol. Biol. *328*, 581–592. https://doi.org/10.1016/S0022-2836(03)00319-X.

Neupert, W., and Herrmann, J.M. (2007). Translocation of Proteins into Mitochondria. Annu. Rev. Biochem. *76*, 723–749. https://doi.org/10.1146/ANNUREV.BIOCHEM.76.052705.163409.

Nguyen, D., Stutz, R., Schorr, S., Lang, S., Pfeffer, S., Freeze, H.H., Förster, F., Helms, V., Dudek, J., and Zimmermann, R. (2018). Proteomics reveals signal peptide features determining the client specificity in human TRAP-dependent ER protein import. Nat. Commun. *9*, 3765. https://doi.org/10.1038/s41467-018-06188-z.

Nguyen, M., Millar, D.G., Yong, V.W., Korsmeyer, S.J., and Shore, G.C. (1993). Targeting of Bcl-2 to the mitochondrial outer membrane by a COOH-terminal signal anchor sequence. J. Biol. Chem. *268*, 25265–25268. https://doi.org/https://doi.org/10.1016/S0021-9258(19)74386-5.

Nguyen Ba, A.N., Pogoutse, A., Provart, N., and Moses, A.M. (2009). NLStradamus: A simple Hidden Markov Model for nuclear localization signal prediction. BMC Bioinformatics *10*. https://doi.org/10.1186/1471-2105-10-202.

Nitta, N., Sugimura, T., Isozaki, A., Mikami, H., Hiraki, K., Sakuma, S., Iino, T., Arai, F., Endo, T., Fujiwaki, Y., et al. (2018). Intelligent Image-Activated Cell Sorting. Cell *175*, 266-276.e13. https://doi.org/10.1016/j.cell.2018.08.028.

Nyathi, Y., Wilkinson, B.M., and Pool, M.R. (2013). Co-translational targeting and translocation of proteins to the endoplasmic reticulum. Biochim. Biophys. Acta - Mol. Cell Res. *1833*, 2392–2402. https://doi.org/10.1016/J.BBAMCR.2013.02.021.

Okreglak, V., and Walter, P. (2014). The conserved AAA-ATPase Msp1 confers organelle specificity to tail-anchored proteins. PNAS *111*, 8019–8024. https://doi.org/10.1073/pnas.1405755111.

Olzscha, H., Schermann, S.M., Woerner, A.C., Pinkert, S., Hecht, M.H., Tartaglia, G.G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F.U., and Vabulas, R.M. (2011). Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions. Cell *144*, 67–78. https://doi.org/10.1016/J.CELL.2010.11.050.

Omura, T. (1998). Mitochondria-targeting sequence, a multi-role sorting sequence recognized at all steps of protein import into mitochondria. J. Biochem. *123*, 1010–1016. https://doi.org/10.1093/OXFORDJOURNALS.JBCHEM.A022036.

Ott, C.M., and Lingappa, V.R. (2002). Integral membrane protein biosynthesis: Why topology is hard to predict. J. Cell Sci. *115*, 2003–2009. https://doi.org/https://doi.org/10.1242/jcs.115.10.2003.

Paetzel, M., Karla, A., Strynadka, N.C.J., and Dalbey, R.E. (2002). Signal peptidases. Chem. Rev. *102*, 4549–4579. https://doi.org/10.1021/CR010166Y.

Park, S.H., Kukushkin, Y., Gupta, R., Chen, T., Konagai, A., Hipp, M.S., Hayer-Hartl, M., and Hartl, F.U. (2013). PolyQ Proteins Interfere with Nuclear Degradation of Cytosolic Proteins by Sequestering the Sis1p Chaperone. Cell *154*, 134–145. https://doi.org/10.1016/J.CELL.2013.06.003.

Pedrazzini, E., Villa, A., Longhi, R., Bulbarelli, A., and Borgese, N. (2000). Mechanism of residence of cytochrome b(5), a tail-anchored protein, in the endoplasmic reticulum. J. Cell Biol. *148*, 899–914. https://doi.org/10.1083/jcb.148.5.899.

Pereira Mendes, M., Hickman, R., Van Verk, M.C., Nieuwendijk, N.M., Reinstädler, A., Panstruga, R., Pieterse, C.M.J., and Van Wees, S.C.M. (2021). A family of pathogen-induced cysteine-rich transmembrane proteins is involved in plant disease resistance. Planta *253*, 1–17. https://doi.org/10.1007/S00425-021-03606-3.

Pidasheva, S., Canaff, L., Simonds, W.F., Marx, S.J., and Hendy, G.N. (2005). Impaired cotranslational processing of the calcium-sensing receptor due to signal peptide missense mutations in familial hypocalciuric hypercalcemia. Hum. Mol. Genet. *14*, 1679–1690. https://doi.org/10.1093/HMG/DDI176.

Pleiner, T., Tomaleri, G.P., Januszyk, K., Inglis, A.J., Hazu, M., and Voorhees, R.M. (2020).

Structural basis for membrane insertion by the human ER membrane protein complex. Science *369*, 433–436. https://doi.org/10.1126/science.abb5008.

Qin, Q., Zhao, T., Zou, W., Shen, K., and Wang, X. (2020). An Endoplasmic Reticulum ATPase Safeguards Endoplasmic Reticulum Identity by Removing Ectopically Localized Mitochondrial Proteins. Cell Rep. *33*, 108363. https://doi.org/10.1016/j.celrep.2020.108363.

Quiroga, R., Trenchi, A., Montoro, A.G., Taubas, J.V., and Maccioni, H.J.F. (2013). Short transmembrane domains with high-volume exoplasmic halves determine retention of type II membrane proteins in the golgi complex. J. Cell Sci. *126*, 5344–5349. https://doi.org/10.1242/jcs.130658.

Rajpar, M.H., Koch, M.J., Davies, R.M., Mellody, K.T., Kielty, C.M., and Dixon, M.J. (2002). Mutation of the signal peptide region of the bicistronic gene DSPP affects translocation to the endoplasmic reticulum and results in defective dentine biomineralization. Hum. Mol. Genet. *11*, 2559–2565. https://doi.org/10.1093/HMG/11.21.2559.

Rao, M., Okreglak, V., Chio, U.S., Cho, H., Walter, P., and Shan, S.O. (2016). Multiple selection filters ensure accurate tail-anchored membrane protein targeting. Elife *5*, e21301. https://doi.org/10.7554/ELIFE.21301.

Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature *450*, 663–669. https://doi.org/10.1038/nature06384.

Ravanelli, S., den Brave, F., and Hoppe, T. (2020). Mitochondrial Quality Control Governed by Ubiquitin. Front. Cell Dev. Biol. *8*. https://doi.org/10.3389/FCELL.2020.00270.

Rayner, J.C., and Pelham, H.R.B. (1997). Transmembrane domain-dependent sorting of proteins to the ER and plasma membrane in yeast. EMBO J. *16*, 1832–1841. https://doi.org/10.1093/emboj/16.8.1832.

Reggiori, F., and Pelham, H.R.B. (2002). A transmembrane ubiquitin ligase required to sort membrane proteins into multivesicular bodies. Nat. Cell Biol. 2002 42 *4*, 117–123. https://doi.org/10.1038/ncb743.

Requião, R.D., Fernandes, L., de Souza, H.J.A., Rossetto, S., Domitrovic, T., and Palhano, F.L. (2017). Protein charge distribution in proteomes and its impact on translation. PLOS Comput. Biol. *13*, e1005549. https://doi.org/10.1371/journal.pcbi.1005549.

Rodrigo-Brenni, M.C., Gutierrez, E., and Hegde, R.S. (2014). Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. Mol. Cell *55*, 227–237. https://doi.org/10.1016/J.MOLCEL.2014.05.025.

Sabherwal, N., Schneider, K.U., Blaschke, R.J., Marchini, A., and Rappold, G. (2004). Impairment of SHOX nuclear localization as a cause for Léri-Weill syndrome. J. Cell Sci. *117*, 3041–3048. https://doi.org/10.1242/JCS.01152.

Schibich, D., Gloge, F., Pöhner, I., Björkholm, P., Wade, R.C., Von Heijne, G., Bukau, B., and Kramer, G. (2016). Global profiling of SRP interaction with nascent polypeptides. Nature *536*, 219–223. https://doi.org/10.1038/nature19070.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

Schraivogel, D., Kuhn, T.M., Rauscher, B., Rodríguez-Martínez, M., Paulsen, M., Owsley, K., Middlebrook, A., Tischer, C., Ramasz, B., Ordoñez-Rueda, D., et al. (2022). High-speed fluorescence image–enabled cell sorting. Science *375*, 315–320. https://doi.org/10.1126/SCIENCE.ABJ3013.

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D., Schwappach, B., and Weissman, J.S. (2008). The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane. Cell *134*, 634–645. https://doi.org/10.1016/j.cell.2008.06.025.

Seppen, J., Steenken, E., Lindhout, D., Bosma, P.J., and Oude Elferink, R.P.J. (1996). A mutation which disrupts the hydrophobic core of the signal peptide of bilirubin UDP-glucuronosyltransferase, an endoplasmic reticulum membrane protein, causes Crigler-Najjar type IIs. FEBS Lett. *390*, 294–298. https://doi.org/10.1016/0014-5793(96)00677-1.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nat. Rev. Genet. *16*, 299–311. https://doi.org/10.1038/nrg3899.

Shaner, N.C., Lambert, G.G., Chammas, A., Ni, Y., Cranfill, P.J., Baird, M.A., Sell, B.R., Allen, J.R., Day, R.N., Israelsson, M., et al. (2013). A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat. Methods *10*, 407. https://doi.org/10.1038/NMETH.2413.

Sharpe, H.J., Stevens, T.J., and Munro, S. (2010). A comprehensive comparison of transmembrane domains reveals organelle-specific properties. Cell *142*, 158–169. https://doi.org/10.1016/j.cell.2010.05.037.

Shurtleff, M.J., Itzhak, D.N., Hussmann, J.A., Schirle Oakdale, N.T., Costa, E.A., Jonikas, M., Weibezahn, J., Popova, K.D., Jan, C.H., Sinitcyn, P., et al. (2018). The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. Elife *7*. https://doi.org/10.7554/eLife.37018.

De Silvestris, M., D'Arrigo, A., and Borgese, N. (1995). The targeting information of the mitochondrial outer membrane isoform of cytochrome *b* 5 is contained within the carboxyl-terminal region. FEBS Lett. *370*, 69–74. https://doi.org/10.1016/0014-5793(95)00797-D.

Smith, J.D., Schlecht, U., Xu, W., Suresh, S., Horecka, J., Proctor, M.J., Aiyar, R.S., Bennett, R.A.O., Chu, A., Li, Y.F., et al. (2017). A method for high-throughput production of sequence-verified DNA libraries and strain collections. Mol. Syst. Biol. *13*, 913. https://doi.org/10.15252/msb.20167233.

Stornaiuolo, M., Lotti, L. V., Borgese, N., Torrisi, M.R., Mottola, G., Martire, G., and Bonatti, S. (2003). KDEL and KKXX Retrieval Signals Appended to the Same Reporter Protein Determine Different Trafficking between Endoplasmic Reticulum, Intermediate Compartment, and Golgi Complex. Mol. Biol. Cell *14*, 889–902. https://doi.org/10.1091/MBC.E02-08-0468.

Swanson, R., Locher, M., and Hochstrasser, M. (2001). A conserved ubiquitin ligase of the nuclear envelope/endoplasmic reticulum that functions in both ER-associated and Matα2 repressor degradation. Genes Dev. *15*, 2660–2674. https://doi.org/10.1101/GAD.933301.

Symons, M., and Rusk, N. (2003). Control of Vesicular Trafficking by Rho GTPases. Curr. Biol. *13*, R409–R418. https://doi.org/10.1016/S0960-9822(03)00324-5.

Tian, S., Wu, Q., Zhou, B., Choi, M.Y., Ding, B., Yang, W., and Correspondence, M.D. (2019). Proteomic Analysis Identifies Membrane Proteins Dependent on the ER Membrane Protein Complex. CellReports *28*, 2517-2526.e5. https://doi.org/10.1016/j.celrep.2019.08.006.

Tong, A.H.Y., and Boone, C. (2007). High-throughput strain construction and systematic synthetic lethal screening in Saccharomyces cerevisiae. Methods Microbiol. *36*, 369–386. https://doi.org/10.1016/S0580-9517(06)36016-3.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364–2368. https://doi.org/10.1126/science.1065810.

Tong, Z., Kim, M.-S., Pandey, A., and Espenshade, P.J. (2014). Identification of Candidate Substrates for the Golgi Tul1 E3 Ligase Using Quantitative diGly Proteomics in Yeast. Mol. Cell. Proteomics *13*, 2871–2882. https://doi.org/10.1074/MCP.M114.040774.

Valdez-Taubas, J., and Pelham, H. (2005). Swf1-dependent palmitoylation of the SNARE Tlg1 prevents its ubiquitination and degradation. EMBO J. *24*, 2524–2532. https://doi.org/10.1038/SJ.EMBOJ.7600724.

Venancio, T.M., and Aravind, L. (2010). CYSTM, a novel cysteine-rich transmembrane module with a role in stress tolerance across eukaryotes. Bioinformatics *26*, 149–152. https://doi.org/10.1093/bioinformatics/btp647.

Vitali, D.G., Sinzel, M., Bulthuis, E.P., Kolb, A., Zabel, S., Mehlhorn, D.G., Figueiredo Costa, B., Farkas, Á., Clancy, A., Schuldiner, M., et al. (2018). The GET pathway can increase the risk of mitochondrial outer membrane proteins to be mistargeted to the ER. J. Cell Sci. *131*, jcs211110. https://doi.org/10.1242/jcs.211110.

Van Der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., and Yu, T. (2014). Scikit-image: Image processing in python. PeerJ *2*, e453. https://doi.org/10.7717/peerj.453.

Walter, P., Ibrahimi, I., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein. J. Cell Biol. *91*, 545–550. https://doi.org/10.1083/JCB.91.2.545.

Wang, C., Lu, T., Emanuel, G., Babcock, H.P., and Zhuang, X. (2019). Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. PNAS *166*, 10842–10851. https://doi.org/10.1073/pnas.1903808116.

Wang, F., Brown, E.C., Mak, G., Zhuang, J., and Denic, V. (2010). A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. Mol. Cell *40*, 159–171. https://doi.org/10.1016/J.MOLCEL.2010.08.038.

Waterham, H.R., Titorenko, V.I., Haima, P., Cregg, J.M., Harder, W., and Veenhuis, M. (1994). The Hansenula polymorpha PER1 gene is essential for peroxisome biogenesis and encodes a

peroxisomal matrix protein with both carboxy- and amino-terminal targeting signals. J. Cell Biol. *127*, 737. https://doi.org/10.1083/JCB.127.3.737.

Weill, U., Yofe, I., Sass, E., Stynen, B., Davidi, D., Natarajan, J., Ben-Menachem, R., Avihou, Z., Goldman, O., Harpaz, N., et al. (2018). Genome-wide SWAp-Tag yeast libraries for proteome exploration. Nat. Methods *15*, 617–622. https://doi.org/10.1038/s41592-018-0044-9.

Weir, N.R., Kamber, R.A., Martenson, J.S., and Denic, V. (2017). The AAA protein Msp1 mediates clearance of excess tail-anchored proteins from the peroxisomal membrane. Elife *6*. https://doi.org/10.7554/eLife.28507.

Williams, R., Peisajovich, S.G., Miller, O.J., Magdassi, S., Tawfik, D.S., and Griffiths, A.D. (2006). Amplification of complex gene libraries by emulsion PCR. Nat. Methods *3*, 545–550. https://doi.org/10.1038/nmeth896.

Woerner, A.C., Frottin, F., Hornburg, D., Feng, L.R., Meissner, F., Patra, M., Tatzelt, J., Mann, M., Winklhofer, K.F., Hartl, F.U., et al. (2016). Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. Science *351*, 173–176. https://doi.org/10.1126/SCIENCE.AAD2033.

Xiao, T., Shakya, V.P., and Hughes, A.L. (2021). ER targeting of non-imported mitochondrial carrier proteins is dependent on the GET pathway. Life Sci. Alliance *4*, e202000918. https://doi.org/10.26508/lsa.202000918.

Xu, Y., Yu, Z., Zhang, D., Huang, J., Wu, C., Yang, G., Yan, K., Zhang, S., and Zheng, C. (2018). CYSTM, a Novel Non-Secreted Cysteine-Rich Peptide Family, Involved in Environmental Stresses in Arabidopsis thaliana. Plant Cell Physiol. *59*, 423–438. https://doi.org/10.1093/pcp/pcx202.

Yan, X., Stuurman, N., Ribeiro, S.A., Tanenbaum, M.E., Horlbeck, M.A., Liem, C.R., Jost, M., Weissman, J.S., and Vale, R.D. (2021). High-content imaging-based pooled CRISPR screens in mammalian cells. J. Cell Biol. *220*, e202008158. https://doi.org/10.1083/JCB.202008158.

Yang, M., Ellenberg, J., Bonifacino, J.S., and Weissman, A.M. (1997). The transmembrane domain of a carboxyl-terminal anchored protein determines localization to the endoplasmic reticulum. J. Biol. Chem. *272*, 1970–1976. https://doi.org/10.1074/jbc.272.3.1970.

Yang, X., Arines, F.M., Zhang, W., and Li, M. (2018). Sorting of a multi-subunit ubiquitin ligase complex in the endolysosome system. Elife *7*, e33116. https://doi.org/10.7554/ELIFE.33116.

Yofe, I., Weill, U., Meurer, M., Chuartzman, S., Zalckvar, E., Goldman, O., Ben-Dor, S., Schütze, C., Wiedemann, N., Knop, M., et al. (2016). One library to make them all: Streamlining the creation of yeast libraries via a SWAp-Tag strategy. Nat. Methods *13*, 371–378. https://doi.org/10.1038/nmeth.3795.

Yorimitsu, T., Sato, K., and Takeuchi, M. (2014). Molecular mechanisms of Sar/Arf GTPases in vesicular trafficking in yeast and plants. Front. Plant Sci. *5*. https://doi.org/10.3389/FPLS.2014.00411.

Zhang, X., and Shan, S.O. (2014). Fidelity of cotranslational protein targeting by the signal recognition particle. Annu. Rev. Biophys. *43*, 381–408. https://doi.org/10.1146/ANNUREV-BIOPHYS-051013-022653.

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., and Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support *11045*. https://doi.org/10.1007/978-3-030-00889-5_1.

Zimmermann, R., Eyrisch, S., Ahmad, M., and Helms, V. (2011). Protein translocation across the ER membrane. Biochim. Biophys. Acta *1808*, 912–924. https://doi.org/10.1016/J.BBAMEM.2010.06.015.

10 APPENDIX II

10.1 ACKNOWLEDGEMENTS

10.2 CURRICULUM VITAE