Strategies and Bottlenecks Towards Humanizing the Proteasome Core in Yeast

Mudabir Abdullah

A Thesis

In the Department of Biology

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Biology) at

Concordia University

Montreal, Quebec, Canada

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Mudabir Abdullah

Entitled: Strategies and Bottlenecks Towards Humanizing Proteasome Core in Yeast

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Dave Mumby

_____ External
Dr. Paul Rowley

_____ External to program
Dr. Paul Joyce

_____ Examiner
Dr. Malcolm Whiteway

_____ Examiner
Dr. Vincent Martin

_____ Thesis supervisor
Dr. Aashiq Kachroo

Approved by

_____
Dr. Graduate Program Director

Date

_____
Dr. Dean

Abstract

Strategies and Bottlenecks Towards Humanizing the Proteasome Core in Yeast

Mudabir Abdullah, Ph.D.
Concordia University, 2023

All organisms are derived from a common ancestor and share several biological processes. The principle of evolutionary conservation of genes between species enables their investigation in simpler model organisms. These evolutionary conserved genes encode essential cellular machinery whose failures are linked to diseases in humans. Despite a billion years of evolutionary divergence, budding yeast shares several thousand protein coding-genes with humans. Recent systematic studies have identified many human orthologs that can individually complement a lethal growth defect conferred by the loss of the corresponding yeast gene. Computational analysis of many properties of orthologous gene pairs revealed functional replaceability is not well-explained by sequence similarity between the human and yeast genes. Instead, it is a property of specific protein complexes and pathways "genetic modularity", that broadly defines the human protein's ability to interact with yeast proteins such that some genetic modules are entirely non-replaceable (e.g., DNA replication initiation complexes, splicing machinery), whereas, some are entirely replaceable, including the proteasome complex, a highly conserved, multi-protein complex comprising ~33 proteins.

The modularity paradigm allows if the entire yeast and human systems are, to a first approximation, interchangeable (at least in yeast). My thesis aims to humanize the yeast proteasome core complex comprising 14 subunits. In Chapter 1, I present a review of concepts relevant to humanized yeast and our recent efforts to use humanized yeast models to study human biology, disease, and evolution. In Chapter 2, I describe efforts to humanize the entire alpha proteasome core in yeast. In

the process, I describe a novel rapid, scalable, and combinatorial genome engineering strategy, by Marker-less Enrichment and Recombination of Genetically Engineered loci (MERGE) in yeast. In Chapter 3, I demonstrate the humanization of the non-replaceable yeast β core subunits revealing the role of species-specific protein-protein interactions and genetic modularity in functional replaceability. Finally, in Chapter 4, I discuss future efforts to humanize the proteasome core in its entirety in yeast, including the assembly chaperones required for optimal assembly of the multi-subunit core.

Acknowledgement

All praises to Allah and His blessing for the completion of this thesis. I thank Allah for all the opportunities, trials and strength that have been showered on me during these five years.

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Aashiq H Kachroo for letting me be part of the Kachroo lab and allowing me the freedom to follow the ideas related to any project that interested me at any given time. This journey was not possible without his constant assistance, encouragement, and help.

I would like to thank my committee members Dr. Malcolm Whiteway and Dr.Vincent Martin for their valuable feedback regarding the project.

I am extremely grateful to Genome foundry members especially Dr. Smita Amarnath for mentorship and helpful discussions.

I extend gracious thanks to Kachroo lab members Sarmin Sultana, Britney Greco, Devina Singh, Farhat Zaffar, Michelle Vandeloo, Aaliya Naaz, Jeffery Bouffard, Homin Jeong, Courtney Gamache, Patrick Kennard, Yu Xing Ren, Negin Neiksirat. Good Times! We were a community. A special thanks to Brittany Greco for making this experience intellectually stimulating, helpful scientific discussions, and always being there in good and tough times. I would also like to thank members of the other labs in GE especially Martin lab for sharing reagents and helpful discussions.

This journey wouldn't have been possible without support of my beloved wife Shafia Mushtaq who stood by me through all my travails and my absence.

Finally, I must express my very profound gratitude to my Family for unconditional support, sisters, nephew, and nieces for all the love and encouragement. A special thanks to my brother Umar Abdullah for providing me with an unfailing support and taking care of entire family in my absence.

Dedications

This thesis is dedicated to the loving memory of my family members who I lost during my PhD. I miss you every day and I wish you were here with us today.

Mohammad Abdullah Kawa

Bilal Ahmed Wani

Mohammad Sultan Wani .

Contribution of authors

The research presented in this thesis was designed by Mudabir Abdullah, Sarmin Sultana and Dr. Aashiq H. Kachroo

Chapter1:

Text from Review Humanized yeast to model human biology, disease, and evolution (*Dis Model Mech* (2022) 15 (6): dmm049309) Equal Contribution

Chapter2:

Mudabir Abdullah, Brittany Greco, Aashiq H Kachroo designed the research, Mudabir Abdullah, Britney Greco, Jon Laurent performed experiments, Michelle Vandeloo analyzed whole genome sequencing data, Rhidhiman K Garge, Daniel R Boutz helped with Mass spectroscopy.

Chapter3:

Sarmin Sultana contributed to variant screening, tail swaps, and complementation assays, Jianhui Li and Mark Hochstrasser performed biochemical assay.

Table of content

List of figures

List of tables

List of abbreviations

CAS	CRISPR-associated.
CFU	Colony Forming Units
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CP	Core Particle
DSB	Double-Strand Break
FOA	Fluoroorotic Acid
HDR	Homology- directed DNA Repair
HR	Homologous Recombination
GFP	Green fluorescent protein
gRNA	guide RNA
GI	Genetic interactions
MERGE	Marker-less Enrichment and Recombination of Genetically Engineered loci
MM	Magic Marker
NGS.	Next Generation Sequencing
PAM	Protospacer adjacent motif
PPI	Protein -protein interactions
PPIs	Protein-Protein Interaction(s)
RP	Regulatory Particle
SGA	Synthetic Genetic Array
TS	Tail-Swap

Chapter 1: Introduction to thesis

1.1 Humanized yeast

The Functional conservation of genes between humans and budding yeast highlights the value of using distant model organism for studying human processes (Dixon et al.., 2008; Dolinski et al.., 2007; Sonnhammer et al.., 2015; Tugendreich et al.., 1994). With over 2000 (~30% of its genome) shared genes with humans (**Fig. 1.1**) and easy genetic manipulation, budding yeast is useful model for large scale testing of orthologous human gene function. The avaialibilty of genome-wide knockout, temperature-sensitive, and titratable promoter yeast collections, and the human ORFeome collection (Lamesch et al., 2007) along with rapid growth and simple conversion between haploid and diploid forms, makes yeast a unique model for systematic humanization assays. Humaniziation involves cloning the orthologous human cDNA in yeast expression vector and transforming in yeast strains harboring conditional knockout alleles of the corresponding yeast genes (Kachroo et al., 2015). If the yeast gene is essential, viable yeast cells that express the human gene alone suggest functional replaceability (**Fig. 1.1A**). Alternatively, CRISPR/Cas9 enables precise and marker less insertion of human coding sequences at the corresponding genomic loci, thus replacing the yeast gene while maintaining the native regulation (Akhmetov et al., 2018; Brzác˘ová et al., 2021). Therefore, humanization assays provide a simple growth-based readout of function while bringing together all the benefits of using yeast cells to study human genetics (Botstein et al.., 2011; Laurent et al.., 2016).

Recent studies have shown that many conserved human genes can substitute for their yeast equivalents and sustain yeast growth, (Garge et al., 2020; Hamza et al., 2015, 2020; Kachroo et al., 2015; Laurent et al., 2020; Sun et al., 2016), creating reagents to directly test human gene function

in tractable system. Humanized yeast also enables testing of the effects of human genetic variation on function at scale and can serve as biological reagent to discover novel therapeutic targets, repurpose drugs and characterize disease mutations (**Fig. 1.1B**).

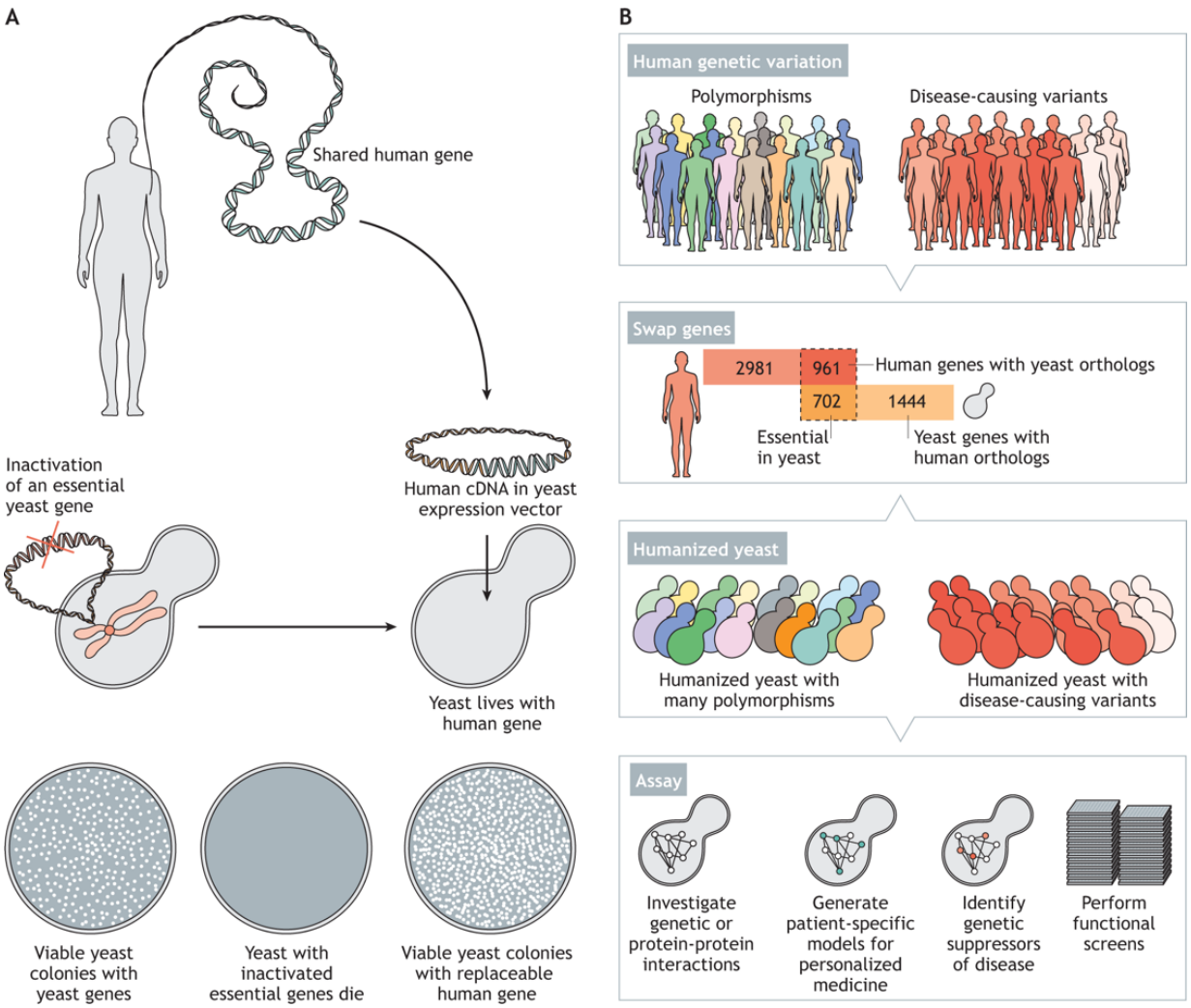


Fig.1.1 Swapping conserved human genes in yeast.

(A) General outline of yeast humanization assays. Shared human protein-coding sequences are cloned in yeast expression vectors. The wild-type yeast that express functional essential genes are viable, as indicated by colonies growing on Petri dishes. The deletion or conditional knockout of the essential yeast gene causes lethality, resulting in an absence of colonies on Petri dishes. However, yeast is viable if the expression of the orthologous human gene can

compensate for the function of the yeast counterpart, despite nearly a billion years of divergent evolution, as indicated by growth on Petri dishes similar to that of wild-type yeast. (B) There is extensive genetic polymorphism in critical human genes, and some of these mutations often lead to disease. Yeast shares 2146 orthologs with humans, of which 702 are essential in yeast. Comparatively, owing to gene amplification, humans share 3942 genes with yeast, of which 961 human orthologs relate to corresponding essential counterparts in yeast [data obtained from Inparanoid (Sonnhammer and Östlund, 2015)]. By functionally replacing the conserved human genes in yeast, the resulting humanized yeast become a tractable system for testing human genetic variation in the context of a simplified cell. These yeast–human gene swaps allow researchers to characterize genetic or protein–protein interactions relevant to disease, build entire human pathways in yeast, generate personalized yeast strains for each unique human variant, identify genetic suppressors of human disease, and provide a platform to identify novel therapeutics. Image concept credit: Andrew Horton.

1.2 Principles governing functional conservation

While functional complementation tests of human genes in yeast have been explored extensively in the past (Balakrishnan et al., 2012; Heinicke et al., 2007), but recent studies have taken a systems approach to test hundreds of shared human genes for functional replaceability in yeast (Garge et al., 2020; Hamza et al., 2015, 2020; Kachroo et al., 2015; Laurent et al., 2020; Sun et al., 2016) (**Fig. 1.2A**). These large-scale have revealed unexpected features that govern functional replaceability, for example, it was found that sequence similarity is not the best predictor of functional replaceability. Instead, the replaceability of human-yeast orthologous pair that have undergone no observable duplication in either lineage (referred to as 1:1 orthologs; **Fig. 1.2A**), is mostly determined at the level of pathways and protein complexes rather than at the level of individual genes, a property known as genetic modularity. The genetic modularity broadly defines the human protein's ability to interact with yeast proteins in a manner like that of the replaced yeast protein (**Fig. 1.2B**). Data showed that some genes that form a functional genetic module were entirely replaceable, such as those encoding components of the proteasome complex and of the sterol and heme biosynthesis pathways. By contrast, other modules, such as the DNA replication initiation complex, were altogether non-replaceable (Kachroo et al., 2015) (**Fig. 1.3C**). The studies also identified several features of shared genes predictive of functional replaceability for human and yeast orthologous pairs that have undergone lineage-specific duplication in the human lineage,

referred to as 1:2 or 1:>2 orthologs (**Fig. 1.2A**), such as relative divergence of human co-orthologs, involvement in GI or PPI networks and subcellular localization (**Fig. 1.2B**) (Laurent et al., 2020).

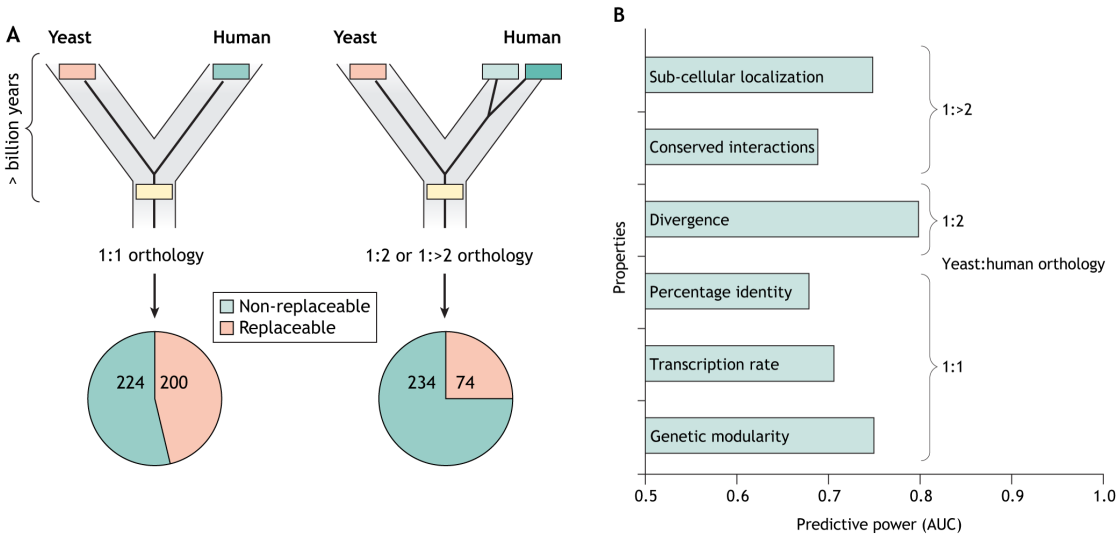


Fig1.2 Systematic humanization of yeast reveals the properties critical for functional replaceability.

(A). Yeast and human genomes share several thousand orthologs that belong to different classes. 1:1 orthologs are shared genes that have acquired no observable duplications in either lineage, whereas 1:2 or 1:>2 refers to orthologs that have undergone duplication in humans (Laurent et al.., 2020). (B) Large-scale replaceability assays identify critical features of shared genes important for functional complementation in yeast. In 1:1 orthologs, genetic modularity is the best predictor of replaceability, followed by transcription rate and amino acid sequence identity. By contrast, in 1:2 or 1:>2 orthologs, the top predictors are divergence of the human or yeast genes, conserved interactions and similar sub-cellular localization. The *x*-axis represents the predictive power calculated as area under the curve (AUC) or receiver operator curve plots [computed from data in Kachroo et al. (2015) and Laurent et al. (2020)].

1.3 Humanizing non-replaceable yeast genes

The concept of ortholog–function conjecture suggests that the orthologs genes from diverse organisms perform equivalent functions (Gabaldón et al., 2013). However, even when orthologous genes perform similar functions, they may not be interchangeable if the organisms have significantly diverged. Nonetheless, functional replaceability can indirectly measure the ability of a human gene to perform most of the essential roles of the yeast counterpart, including critical genetic interactions or protein-protein interactions. Restoring these essential interactions may

facilitate the humanization of non-replaceable yeast genes (**Fig. 1.3A**). However, several factors, including evolution of partners (Kachroo et al., 2015; Teufel et al., 2019), distinct genetic interactions or protein-protein interactions, and non-native expression (Garge et al., 2020), can explain why conserved human genes are not replaceable in yeast. Mutating the non-replaceable human genes can make them replaceable in yeast. Therefore, it is crucial to identify and characterize mutations in human or yeast genes that allow the assimilation of human genes it may not be possible to swap them if the organisms have widely diverged.

Most humanization assays have only tested the functional replaceability of individual genes, but genes do not work in isolation. Investigating whether entire yeast and human systems are interchangeable can enable the humanization of yeast biological processes in their entirety (**Fig. 1.3B, C**). For instance, genetic modules such as the proteasome complex and the cholesterol and heme biosynthesis pathways can be fully humanized because these modules are mainly composed of genes that are individually replaceable in yeast (Kachroo et al., 2015, 2017) (**Fig. 1.3C**). This innovative genetic modularity-based approach could offer a unique opportunity to study polygenic diseases associated with variants belonging to the same protein complex or biosynthetic pathway. These fully humanized yeast strains carrying multiple human genes encoding components of the same biological complex or process will be more suitable for screening human disease-associated alleles as they maintain human-like local genetic interactions or protein-protein interactions. The previous successful attempts, such as the humanization of core histones and the purine biosynthesis pathway in yeast (Agmon et al., 2020; Truong et al., 2017), indicate that this endeavor is feasible.

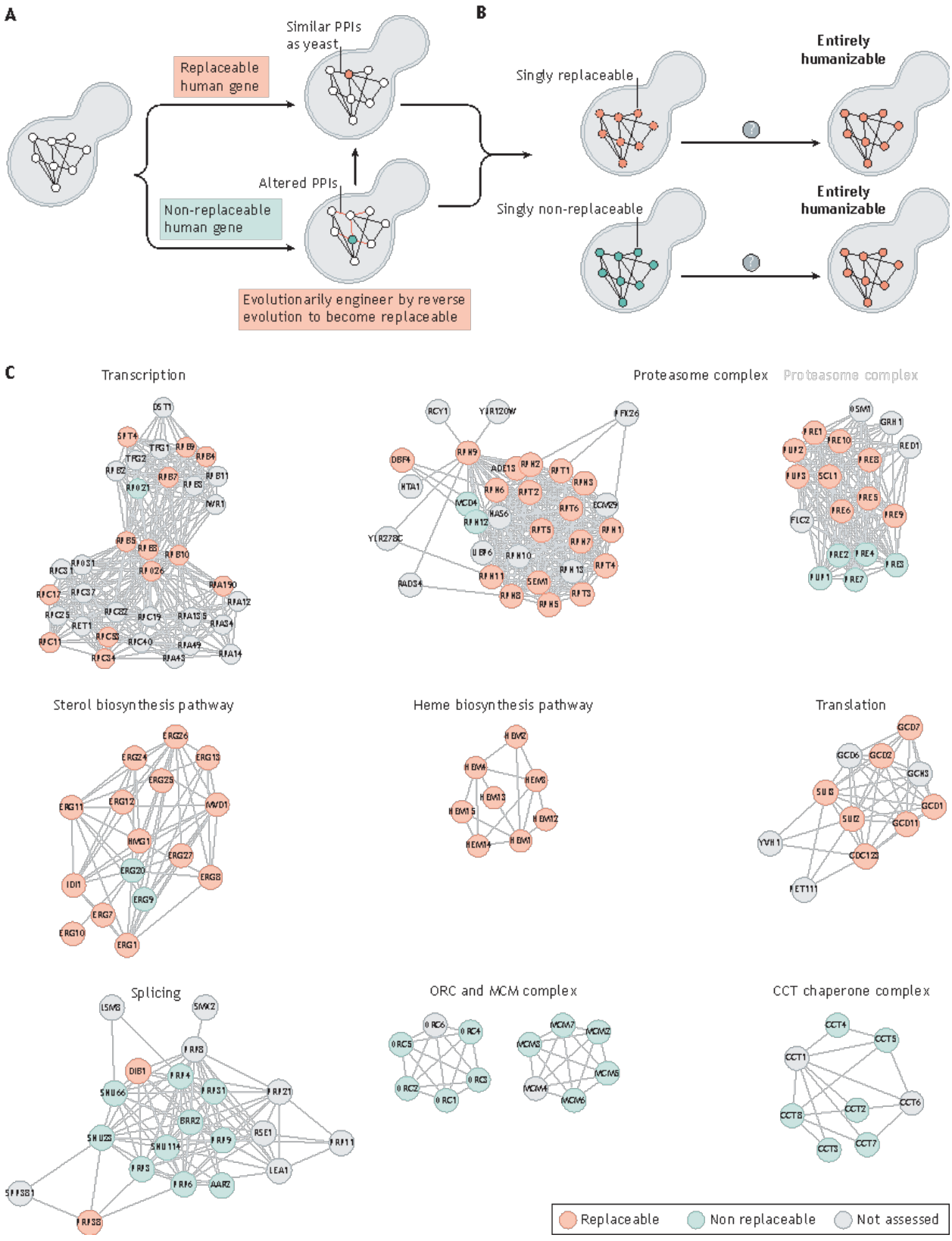

A
Similar PPIs as yeast
Replaceable human gene
Altered PPIs
Non-replaceable human gene
Evolutionarily engineer by reverse evolution to become replaceable
B
Singly replaceable
Entirely humanizable
Singly non-replaceable
Entirely humanizable
C
Transcription
Proteasome complex
Proteasome complex
Sterol biosynthesis pathway
Heme biosynthesis pathway
Translation
Splicing
ORC and MCM complex
CCT chaperone complex
Replaceable
Non replaceable
Not assessed

Fig.1.3. Genetic modules govern functional replaceability.

(A) Lack of functional complementation by a shared human gene in yeast could be attributed to the inability of a human gene to perform critical genetic interactions or protein–protein interactions (PPIs) in yeast. Using a reverse evolution approach and modifying a non-replaceable human gene to complement the yeast ortholog should allow the discovery of critical interactions or other factors, such as diverged mechanisms or regulation between humans and yeast. (B) Genetic modularity is a feature that strongly predicts replaceability and allows researchers to test whether higher-order humanizations of yeast are possible. Some modules can be humanized because most of the individual genes within the module are replaceable, either sequentially or by expressing all humanized components simultaneously. However, non-replaceable modules represent a major challenge and could be humanized if the entire yeast genetic module is replaced simultaneously. (C) Several yeast genes are functionally replaceable by their human equivalents one gene at a time, but many are not. For example, genes encoding components of the transcription and translation machinery, the proteasome complex, and the sterol and heme biosynthesis pathways are mostly replaceable. By contrast, modules such as the splicing complex, the origin recognition complex (ORC), minichromosomal maintenance (MCM) complex and the chaperone-containing TCP-1 (CCT) complex are largely non-replaceable. This distribution of replaceable or non-replaceable human genes in pathways or complexes suggests that these yeast processes are likely humanizable in their entirety, even when individual genes are non-replaceable. The module maps were generated using Cytoscape (version 3.9.1) (Shannon et al.., 2003) with data from Garge et al. (2020), Kachroo et al. (2015) and Laurent et al. (2020) and are meant to illustrate the broad spectrum of functional replaceability across different cellular processes.

1.4 Proteasome complex

Proteins and membranes can become damaged during cellular process, and cells have two proteolytic systems to eliminate these damaged components and maintain protein and organelle turnover. These systems are the proteasome and lysosomal systems. The proteasome degrades short-lived and abnormal proteins with the help of ubiquitin system (Wang et al., 2014). This highly conserved molecular machine is responsible for recycling the cellular proteome (Maupin et al., 2012) and plays a crucial role in many cellular processes, including cell cycle progression, gene expression, transcription and cell signaling. The proteasome works with ubiquitination, recognizing proteins to be degraded by their poly-ubiquitin chain markers. The 26S proteasome consists of the 20S proteasome catalytic “core particle (CP)” and the 19S “regulatory particle (RP)”. Throughout the kingdom of life, CP consists of four stacked heptameric rings. In Archaea and bacteria, CP constitutes single α and β subunits. In eukaryotes, the 26S proteasome is a cylindrical complex

containing a CP of four stacked rings forming a barrel-shaped central catalytic core. Each ring is composed of seven individual proteins. The inner two rings are made of seven β subunits that comprise six protease active sites. These sites are located on the interior surface of the rings such that the target proteins that enter the central pore are degraded. The outer two α-rings each contain seven α subunits whose function is to regulate a "gate" through which proteins enter the catalytic core (Beck, F et al., 2012). The proteasome's function is to degrade defective proteins in a regulated fashion by cleaving them into short peptides via the proteolytic activity of three β subunits (Forster et al., 2013). In humans and yeast, five (β1, β2, β5, β6, β7) of the seven β subunits are initially synthesized with pro-peptides which are cleaved immediately after the correct assembly of the functional proteasome (Chen,P et al., 1996) . The β1, β2, and β5 subunits show different proteolytic activities, such as trypsin-, caspase- and chymotrypsin-like respectively. The RP is composed of lid and base. The RP base subunits are ATPases that play an essential role in the regulation of gating of the α-ring by promoting translocation of unfolded protein substrate into CP. Given the proteasome complex's critical nature, eukaryotes have evolved a highly regulated process to assemble this molecular machine inside cell (Murata et al., 2009). The correct proteasome assembly is achieved by several chaperones that aid in the recruitment, sequential incorporation, and maturation of catalytic β subunits to make a functional proteasome.

Humans possess duplicated copies of some core subunits (2 copies each for α4, β1, β2 and 3 for β5 subunits). Therefore, different kinds of tissue-specific proteasomes are made owing to the type of subunits incorporated in the newly assembled proteasome (Demartino et al., 2007and Sutoh, Y et al., 2012). Constitutive-proteasome, present in most human tissues, comprises catalytic subunits termed β1c (caspase-like, cleaving preferentially after acidic residues), β2c (trypsin-like, cleaving preferentially after basic residues) and β5c (chymotrypsin-like, cleaving preferentially after

hydrophobic residues). However, in human lymphoid tissues and cells that are exposed to the inflammatory cytokine's interferon-γ (IFN-γ) and tumor necrosis factor-α (TNF-α), cells make another variant of the proteasome, called immuno-proteasome, in which β1c, β2c and β5c are replaced by their duplicated versions β1i (branched amino acid preferring, BrAAP), β2i and β5i, respectively. Humans also have one thymus specific β5t subunit, which is exclusively expressed in cortical thymic cells, playing a critical role in the development of T-cells in the thymus that has reduced chymotrypsin-like activity helping in the production of unique peptides in the thymus (Murata, S. et al., 2007).

Proteasome is largely replaceable in yeast by human ortholog gene by plasmid-based complementation (Kachroo et al., 2015). Proteasome complex is an ideal platform to test whether higher order humanization is possible and explore different strategies to humanize non-replaceable subunits. The interaction context with neighboring proteins is a major factor determining replacement of an ortholog. That is, if a significant fraction of a gene's physical or functional interaction partners are replaceable, the gene in question is more likely to be replaceable. As discussed above this phenomenon is called genetic modularity, whereby a protein evolves and retains the ability to successfully replace and assimilate in a genetic module, with the proteins it must interact with to carry out its function. This observation strongly suggests that it may be possible to extend the scope of humanization beyond individual replacements to humanize multiple genes simultaneously, likely starting with those that function in the same complex or biochemical pathway. In Chapter 2, we demonstrated that nearly all subunits in the yeast proteasome are individually replaceable at the native loci and used sequential strategy to humanize the rest subunits to achieve whole-complex humanization. In Chapter 3, we explored the genetic modularity theme

for functional replaceability to assess whether individually non-replaceable proteins can become replaceable in the presence of their interaction partners.

Chapter 2

Rapid, scalable, and combinatorial genome engineering by Marker less Enrichment and Recombination of Genetically Engineered loci (MERGE) in yeast

Mudabir Abdullah, Brittany M. Greco, Jon M. Laurent, Riddhiman K. Garge, Daniel R. Boutz, Michelle Vandeloo, Edward M. Marcotte, Aashiq H. Kachroo

bioRxiv.2022. Version1, Doi: https://doi.org/10.1101/2022.06.17.496490; version 2, SSRN, Doi: http://dx.doi.org/10.2139/ssrn.4208348

Abstract:

A major challenge to rationally building multi-gene processes in yeast arises due to the combinatorics of combining all the individual edits into the same strain. Here, we present a scalable, precise, multi-site genome editing approach that combines all edits without selection markers, using CRISPR-Cas9 and gene drives. We show that engineered loci become resistant to the corresponding CRISPR reagents, allowing the enrichment of distinct genotypes. We demonstrate a highly efficient gene drive that selectively eliminates specific loci by integrating CRISPR-Cas9-mediated Double-Strand Break (DSB) generation and homology-directed recombination with yeast sexual assortment. The method enables Marker-less Enrichment and the Recombination of Genetically Engineered loci (MERGE). We show that MERGE converts single heterologous loci to homozygous loci at ~100% efficiency, independent of chromosomal location. Furthermore, MERGE is equally efficient at converting and combining multiple loci, thus identifying viable intermediate genotypes. Finally, we establish the feasibility of MERGE by engineering a fungal carotenoid biosynthesis pathway and 6 of 7 human α proteasome core proteins into yeast. MERGE,

therefore, lays the foundation for scalable and combinatorial genome editing, enabling pathway engineering and functional genomics in yeast.

2.1 Introduction:

Baker's yeast has long served as a convenient chassis for bioengineering owing to its genetic tractability, versatile metabolism, and ease of culture in the lab. Decades of fundamental research, together with the development of high throughput toolkits and genome engineering capacities, have established yeast as an ideal model eukaryote for system genetics and synthetic biology (Botstein et al., 2011). The availability of many selectable genetic markers and simple conversion between haploid and diploid forms provide avenues to easily combine pairs of genetically engineered loci to understand gene-gene interactions at a global scale (Costanzo et al., 2016). For more extensive genetic alterations, yeast's highly efficient Homologous Recombination (HR) pathway even enables the synthesis of entire chromosomes, although this approach requires iterative use of selection markers and tedious repetitive procedures (Richardson et al., 2017). Nonetheless, the ability to alter large contiguous segments of genomic loci has many applications, such as genome minimization (Suzuki et al., 2015), multiplex genome editing (Barbieri et al., 2017), and the total synthesis of the Mycoplasma and *E.coli* genomes using yeast HR (Gibson et al., 2008; Fredens et al., 2019). In spite of the progress in whole genome engineering, editing intermediate numbers of independent genomic loci–greater than two and in discontiguous regions of the genome still presents a significant technological challenge. While strategies exist for *E. coli* (Lajoie et al., 2013) (e.g., MAGE), diploid organisms such as yeast present the additional challenges for example editing multiple alleles for each (independently assorting) genomic locus. Moreover, while high throughput cloning strategies and the reduced cost of de novo DNA synthesis now allow swapping entire

heterologous pathways or protein complexes into yeast, such efforts frequently require the deletion of corresponding yeast loci, such as the case for efforts to systematically humanize yeast genes (Truong et al., 2017; Agmon et al., 2020), which may entail replacing genes at their corresponding genomic loci to retain native regulation. Beyond these aspects, the expression of each new gene often reveals incompatibilities associated with the engineered pathway. Thus, there is a need for rapid, multi-site, progressive genome editing strategies to address these issues. The efficiency and speed of CRISPR-Cas9-based genome engineering allow straightforward editing of multiple yeast loci in a single strain, eliminating the need for markers (Ran et al., 2013; DiCarlo et al., 2013; Akhmetov et al., 2018; Kachroo et al., 2017; Lee et al., 2015). However, the approach gets progressively more challenging for multi-gene systems when the fitness of the intermediate genotypes is unknown (Figure S2.1). Therefore, rationally building heterologous genetic modules in a yeast surrogate requires a highly scalable and combinatorial genome editing technology. Synthetic Genetic Array (SGA) analysis permits the combination of loci in a fitness-driven manner, requiring markers linked to the modified loci and haploid-specific selections (Kuzmin et al., 2016). Using an SGA-like strategy to build a heterologous multi-gene system would need several unique markers linked to each gene. However, the lack of adequate selection markers limits its application. The Green Monster (GM) method bypasses the marker dependency by using Green Fluorescent Protein (*GFP*) expression as a readout to combine engineered loci (Suzuki et al., 2011). However, implementing the GM strategy can be challenging for heterologous pathway engineering with an added burden of expressing fluorescent cassettes (Labunskyy et al., 2014). Thus, accomplishing large-scale combinatorial genome editing in yeast necessitates new genetic tools that circumvent these requirements. In this work, we describe a novel, CRISPR-Cas9-based method to readily combine genetically engineered loci without the need for markers and exogenous repair templates.

The approach involves generating CRISPR-Cas9-mediated Double-Strand Breaks (DSBs) and highly efficient Homology Directed Repair (HDR) along with yeast mating and sporulation to randomly assort edited sites performing multi-site and marker-less combinations of engineered loci. This selective and successive elimination of specific yeast loci mimics a gene drive (Xu et al., 2020; Yan & Finnigan, 2018; DiCarlo et al., 2015; Roggenkamp et al., 2017) (Figure S2.2) and facilitates Marker-less Enrichment and Recombination of Genetically Engineered yeast loci (MERGE). We show that CRISPR-mediated selection operates like classical selection markers, enabling MERGE to efficiently explore many combinations of genetically engineered loci, revealing a fitness-driven path to engineering any heterologous system in yeast. We further demonstrate that MERGE enables rapid assembly of an entire carotenoid biosynthesis pathway by performing a multi-site and marker-less combination of distinct engineered loci. Finally, using a multiplexed version of MERGE, we humanize a near-complete α proteasome core (6 of 7 subunits) in yeast while revealing a fitness-driven path to the humanization of complex processes.

2.2 Results:

2.2.1 CRISPR-Cas9 allows marker-less selection and enrichment of unique genotypes in yeast

CRISPR-Cas9 enables precise and marker-less editing of both essential and non-essential yeast loci owing to significantly lower error-prone Non-Homologous End Joining (NHEJ) in yeast relative to HDR (Kramer et al., 1994). Therefore, Cas9-sgRNA-induced lethality serves as a rapid test for a functional CRISPR reagent in yeast (Colony Forming Units Observed or $CFU_O = \sim 0$) compared to the Cas9 alone (Colony Forming Units Expected or CFU_E). To verify the ON target activity of CRISPR reagents, the transformation of a plasmid carrying an expression cassette for both Cas9 and a sgRNA targeting a specific locus (pCas9-sgRNA^{locus}) in a yeast strain harboring a

corresponding engineered locus resistant to further targeting allows the survival of colonies similar to the pCas9 alone (CFU_O/CFU_E = ~1) (**Figure 2.1A**).

We tested this strategy by humanizing *α* proteasome core genes in yeast that are functionally replaceable by their human counterparts (Kachroo et al., 2015). Using CRISPR-Cas9, we replaced each yeast *α* proteasome genes with its human ortholog at the native loci (Figure S2.3). Additionally, we modified non-essential loci to test if the CRISPR-mediated selection is broadly applicable and found that all engineered strains are resistant to the corresponding CRISPR-Cas9-sgRNA mediated lethality (CFU_O/CFU_E = ~1, **Figure 2.1B**, Figure S2.4A & S2.4B).

To verify if the resistance of humanized strains is exclusive to its corresponding CRISPR reagent, all single-humanized *α* proteasome strains were mixed in culture (**Figure 2.1C**). To quantify the enrichment of unique genotypes, we inoculated a *ade2Δ::kanMX* haploid strain in the mixture. The pCas9-sgRNA^{ADE2} exclusively enriched for resistant *ade2Δ* genotype, whereas all other genotypes in the mix harboring a wild type *ADE2* locus are inviable (**Figure 2.1C-ii**).

Conversely, the transformation of pCas9-sgRNA^{RPT5} targeting *RPT5* (a base subunit of the proteasome complex), for which all the strains in the mix harbor a wild-type copy, shows no survivors (**Figure 2.1C-iii**). We demonstrate that each CRISPR reagent targeting yeast *α* proteasome genes selected a corresponding humanized genotype from a mix, respectively (**Figure 2.1C-iv**). Thus, CRISPR-Cas9-mediated resistance functions similarly to conventional antibiotic or auxotrophic markers in yeast.

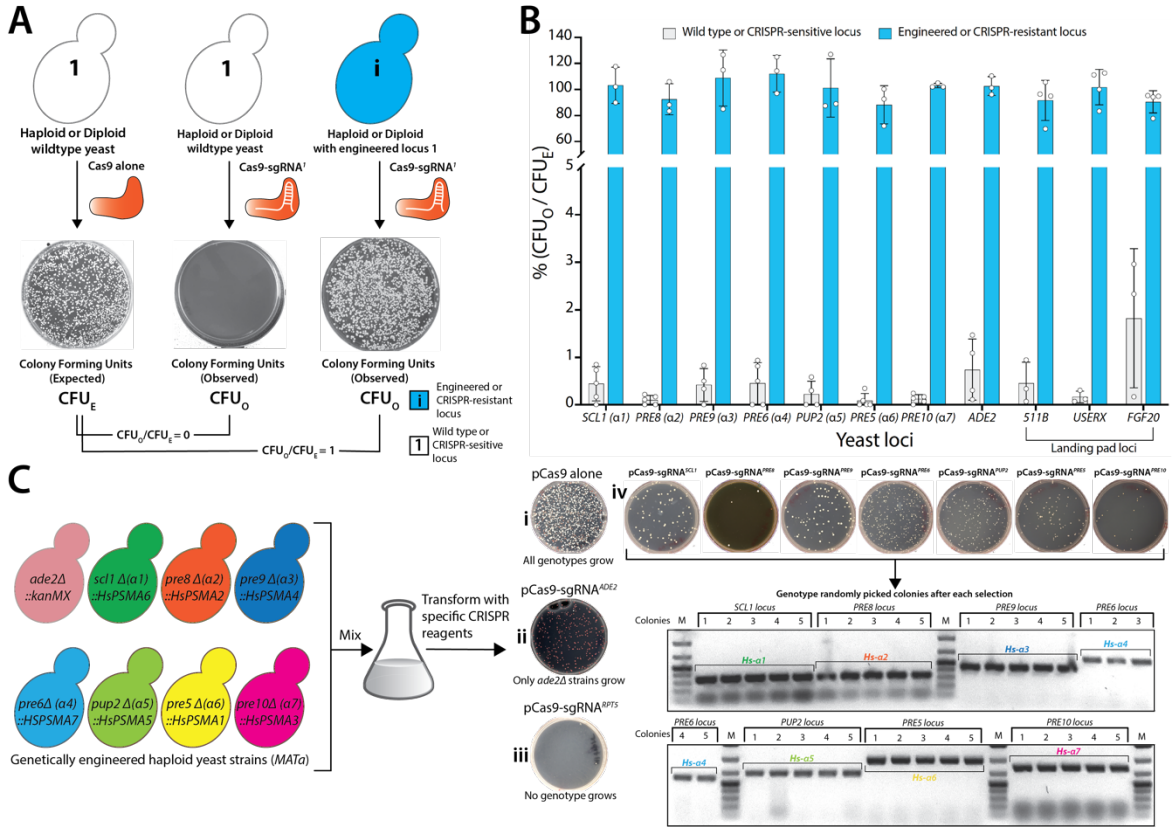


Figure 2.1 CRISPR-mediated selection enables the enrichment of unique genotypes.

(A) pCas9-sgRNA^{1} targeted to any yeast locus (locus 1) leads to lethality (CFU_O) compared to the vector without sgRNA (pCas9 alone; CFU_E). However, the modification of locus 1 to i prevents further targeting by the corresponding CRISPR reagent (CFU_O). A simple readout of CFU_O/CFU_E enables the identification of modified genotypes. (B) Each CRISPR reagent targeted to several wild-type yeast loci shows near 100% lethality (Gray bars or CRISPR-sensitive loci). However, after editing the loci (by generating single-humanized yeast α proteasome genes, *ade2Δ::kanMX* and inserting carotenoid genes at landing pad loci), the strains show resistance to the corresponding CRISPR reagent (Blue bars or CRISPR-resistant loci, CFU_O/CFU_E = ~1). (C) Each unique genetically modified strain can be enriched from the mixture using the corresponding CRISPR reagent. The transformation of pCas9 alone serves as a control (i). Haploid *ade2Δ::kanMX* strain quantifies the efficiency of the CRISPR selection (pCas9-sgRNA^{ADE2}), enriching only the *ade2Δ* homozygous genotypes (red-colony phenotype) (ii). The transformation of pCas9-sgRNA^{RPT5} causes lethality as all the strains in the mix harbor a wild-type copy of the *RPT5* gene (iii). Similarly, each CRISPR reagent specific to individual yeast *α* proteasome genes also selected the corresponding humanized strains as demonstrated by PCR-based genotyping of randomly picked colonies (iv).

2.2.2 MERGE^{0} is nearly 100% efficient at converting loci irrespective of the yeast gene location on the chromosome

Mating of engineered haploid strains with wild type generates a heterozygous genotype. CRISPR-Cas9 reagents targeted to either allele should enable the conversion to homozygous diploids at high efficiency while simultaneously enriching the desired genotype (MERGE level ^{0} or MERGE^{0}) (**Figure 2.2A**). We used *ADE2/ade2Δ::kanMX* heterozygous knockout diploid (hetKO) strain to quantify the efficiency of MERGE^{0}. The wild type *ADE2* allele is susceptible to pCas9-sgRNA^{ADE2} mediated DSB. In contrast, the *ade2Δ::kanMX* allele is resistant, proving a readout of conversion as the loss-of-function of *ADE2* results in a red color colony phenotype. The transformation of both wild-type haploid or diploid yeast with pCas9-sgRNA^{ADE2} showed a lethal phenotype (~0-20 CFU_O on average) (**Figure 2.2B**). However, pCas9-sgRNA^{ADE2} transformation in *ADE2/ade2Δ::kanMX* hetKO strain shows resistance while also simultaneously converting the locus to *Δade2::kanMX* at ~100% efficiency (**Figures 2.2B** & S2.5A). MERGE^{0} similarly performs with comparable proficiencies across multiple loci at single-humanized *a* proteasome and landing pad loci (**Figures 2.2C**, S2.5B & S2.5C). To test if MERGE^{0} can convert any heterozygous yeast locus independent of the position on a chromosome, we explored the strategy across many yeast genes located on chromosome 1. We designed a CRISPR reagent (pCAS9- sgRNA^{kanMX}) to target the het KO diploid yeast strains harboring the *kanMX* cassette instead of a yeast gene. MERGE^{0} converted all heterozygous loci to homozygous wild-type alleles, respectively, with simultaneous loss of *kanMX* cassette (except for *CNE1*) (**Figure 2.2D**, S2.6A & S2.6B). Furthermore, using CRISPR-Cas9 to target a wild-type yeast locus instead of the *kanMX* cassette in a hetKO diploid strains, MERGE^{0} enables single-step gene essentiality screening in yeast, significantly increasing efficiency compared to the classical methods such as Tetrad dissection or SGA analysis (Winzeler et al., 1999; Pan et al., 2004). Except for a non-essential *PRE9 (α3)* gene, all six essential α subunit proteasome genes are inviable post-transformation of the corresponding CRISPR reagents (**Figure 2.2E**).

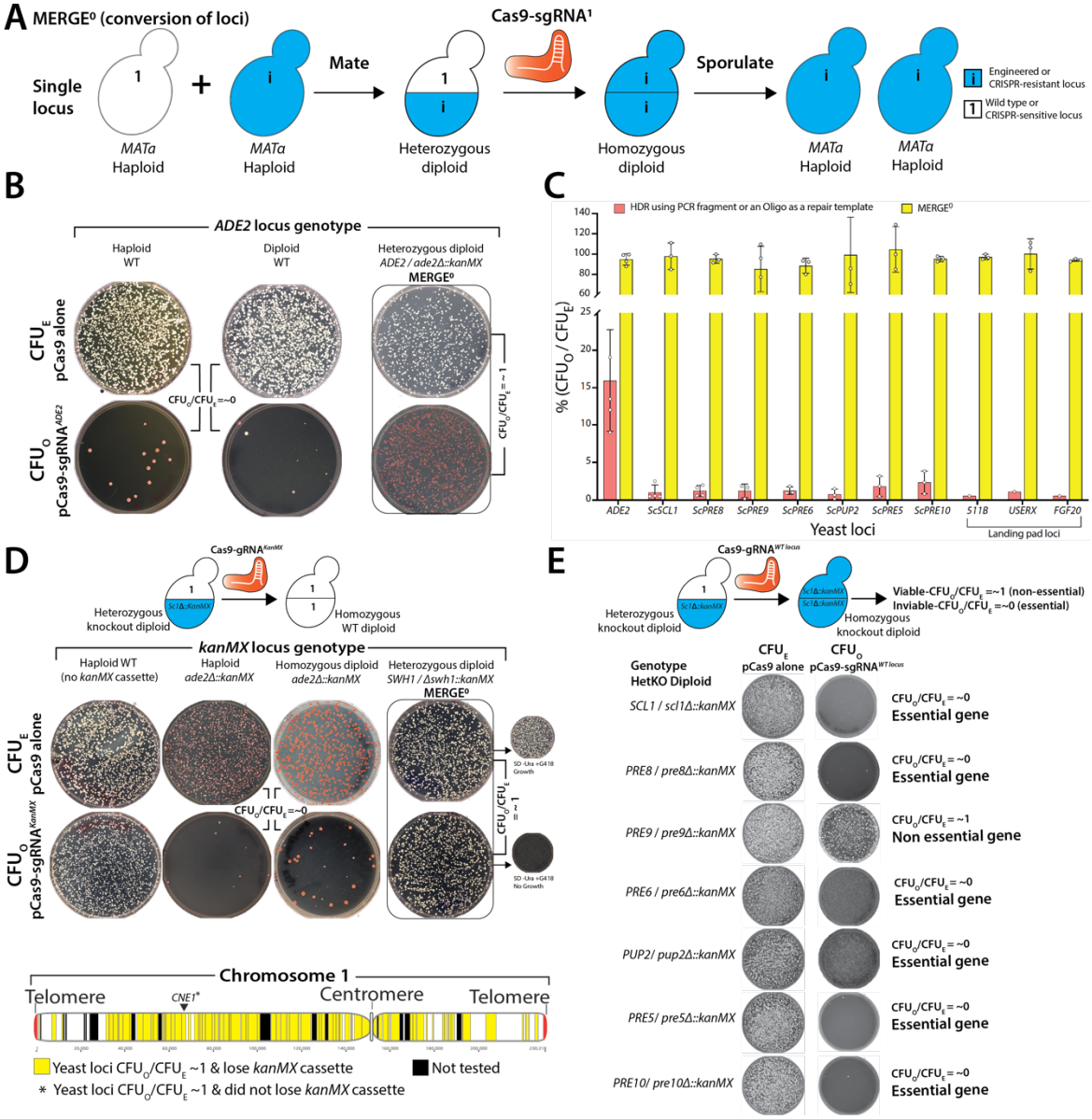


Figure 2.2 MERGE level 0 (MERGE^0) efficiently converts a single heterozygous to a homozygous locus.

(A) Mating haploid yeast strains, each harboring a different allele at a single locus, enables the combination as a heterozygote. CRISPR reagent targeted to one of the alleles (CRISPR-sensitive locus) initiates recombination using the homologous chromosome with a CRISPR-resistant locus as a repair template. Sporulation of the resulting homozygous diploid yields the desired genotype in both mating types. (B) CRISPR reagent (pCas9-sgRNA^{ADE2}) targeted to the wild type *ADE2* locus quantifies the efficiency of MERGE^0. The transformation of pCas9-sgRNA^{ADE2} in the wild type haploid or diploid strains is lethal, with few surviving red colonies suggesting efficient ON-target activity (CFU_O/CFU_E = ~0). However, the transformation of pCas9-sgRNA^{ADE2} in heterozygous

diploid *ADE2/ade2Δ::kanMX*strain shows no lethality ($CFU_O/CFU_E = \sim 1$). Instead, all surviving red colonies suggest efficient conversion to the knockout locus. (C) Similarly, CRISPR reagents targeted to several wild-type yeast loci show near 100% lethality with few surviving colonies while simultaneously providing exogenous repair templates (oligo for *ADE2* locus and PCR fragments for the remaining loci) (Red bars). However, each heterozygous diploid displays resistance to CRISPR reagents respectively ($CFU_O/CFU_E = \sim 1$, Yellow bars). PCR confirmation verified the conversion to the CRISPR-resistant locus. (D) CRISPR reagent targeting *kanMX* cassette (pCas9-sgRNA^{kanMX}) shows no OFF-target activity in a strain lacking the cassette ($CFU_O/CFU_E = \sim 1$). However, haploid or diploid strains harboring *kanMX* cassettes as the only allele show lethality ($CFU_O/CFU_E = \sim 0$), suggesting ON-target activity. However, diploid strain heterozygous for *kanMX* allele (*SWH1*/*swh1Δ::kanMX*) is viable ($CFU_O/CFU_E = \sim 1$), suggesting conversion to the wild type allele as demonstrated by the loss of G418 resistance after performing MERGE^0. Using heterozygous diploid knockout strains (hetKO) arrayed across the entire yeast chromosome I, MERGE^0 similarly converted every knockout allele to the wild type locus (Yellow = $CFU_O/CFU_E = \sim 1$), except in the case of *CNE1* that showed resistance to the CRISPR reagent but did not lose the *kanMX* cassette. (E) Alternatively, using hetKO strains and targeting CRISPR reagent to the wild-type yeast locus enables a single-step gene essentiality assay in yeast. The remarkably high efficiency of MERGE^0 converts every heterozygous yeast locus to homozygous null, therefore, allowing viability only if the gene is non-essential (as in the case of yeast proteasome α3 gene). All other essential loci show lethality after MERGE^0.

2.2.3 MERGE^1 permits a fitness-driven combination of the engineered loci

Given the high efficiency of MERGE^0 at converting a single yeast locus, we sought to test if the method works similarly to convert and combine two distinct loci (MERGE^1). Mating haploid yeast strains, each with one modified locus, facilitates the combination as heterozygotes at two separate loci. CRISPR-Cas9 with two sgRNAs each uniquely targeting their corresponding locus allows for simultaneous generation of two CRISPR-resistant loci (**Figure 2.3A**). To test the pipeline across all humanized α proteasome strains, we first used MERGE^0 to move all singly humanized α proteasome loci from a BY4741 background to an SGA strain background in both mating-types (Figure S2.4B). To verify if MERGE^1 can simultaneously convert two distinct loci, we tested the efficiency of combining paired-humanized *α3/α7* and *α1/α7* genotypes. We found that MERGE^1 is highly efficient as each heterozygous genotype is resistant to the double-sgRNA CRISPR-Cas9 reagent while converting and combining two humanized loci (**Figure 2.3B**, S2.7A). The CRISPR reagent is lethal in wild-type and singly humanized strains (**Figure 2.3B** & S2.7B). The sporulation of viable diploids yields haploid double-humanized strains in both mating-types. Genotyping randomly picked colonies confirmed the conversion to humanized loci (**Figure 2.3B**). Additionally,

the double-humanized *α7/α1* genotype, while viable, shows a sporulation defect (a phenotype associated with Hs*α*7) (Figure S2.7A). Thus, MERGE[1] allows the survival and enrichment of only viable homozygous double-humanized loci without the requirement of any diploid-specific selection. Interestingly, we found that a double-humanized *α5/α7* strain is inviable as a combined genotype, as evidenced by double-sgRNA CRISPR-Cas9-mediated lethality (CFU_O/CFU_E = ~0). PCR-based genotyping of a surviving colony shows homozygosity of only the Hs*α*7 locus (**Figure 2.3C**). In contrast, while a single-humanized Hs*α*6 strain is temperature-sensitive (ts) at 37°C, combination with the neighboring Hs*α*7 gene rescues the ts phenotype (Figure S2.7A′). However, the ts phenotype is associated with one of the variants of Hs*a*6 (Variant 1, 37>Glycine) (Figure S2.8A), whereas another common variant (Variant 2, 37>Valine) shows no growth defect at 37°C (Figure S2.8B). Thus, MERGE[1] revealed the fitness of paired-humanized genotypes and epistasis similar to synthetic genetic interactions without the need for linked markers or haploid-specific selections (Kuzmin et al., 2016).

To systematically determine if there are specific pairwise restrictions to the humanization of *α* subunits, we mated all haploid single-humanized strains obtaining diploid heterozygous genotypes (21 different genotypes). The corresponding CRISPR-Cas9-based selection and a simple readout of CFU_O/CFU_E identified the permitted double-humanized genotypes (17/21 genotypes). We found that only specific double-humanized genotypes are viable, whereas some are not (**Figure 2.3D** & S2.7C). The incompatibility of paired genotypes comprising *Hsα1/α6, Hsα4/α6, Hsα5/α6,* and *Hsα5/α7* may likely be due to the missing neighboring interactions within the α proteasome core, except in the case of *Hsα5/α6* pair (Figure S2.7C′). With a sequential editing strategy, we successfully engineered several paired-humanized genotypes to confirm that MERGE[1] represents

viable/fit paired genotypes. However, the sequential strategy did not provide a clear perspective of incompatibilities (Figure S2.7D).

Given the success of MERGE[1], we examined the strategy to select double-humanized genotypes randomly from a mixture (Mix, Mate & MERGE or MERGE$^{M\&M}$) (**Figure 2.3E**). We inoculated single-humanized α1, α3 and α7 haploid strains of both mating-types as a mix, allowing random mating. Each double-sgRNA CRISPR-Cas9 selection enriched the corresponding paired genotype from the mated mix while converting the wild-type yeast to humanized loci (**Figure2.3E iii & 2.3E-iv**). In mixed culture, the transformation of pCas9-sgRNA$^{Sc\text{-}\alpha1,\alpha6}$, a selection for a non-existing genotype, does not yield any viable genotype (**Figure 2.3F-ii**). Therefore, MERGE$^{M\&M}$ can be scaled to obtain several paired genotype combinations of engineered loci from a mix.

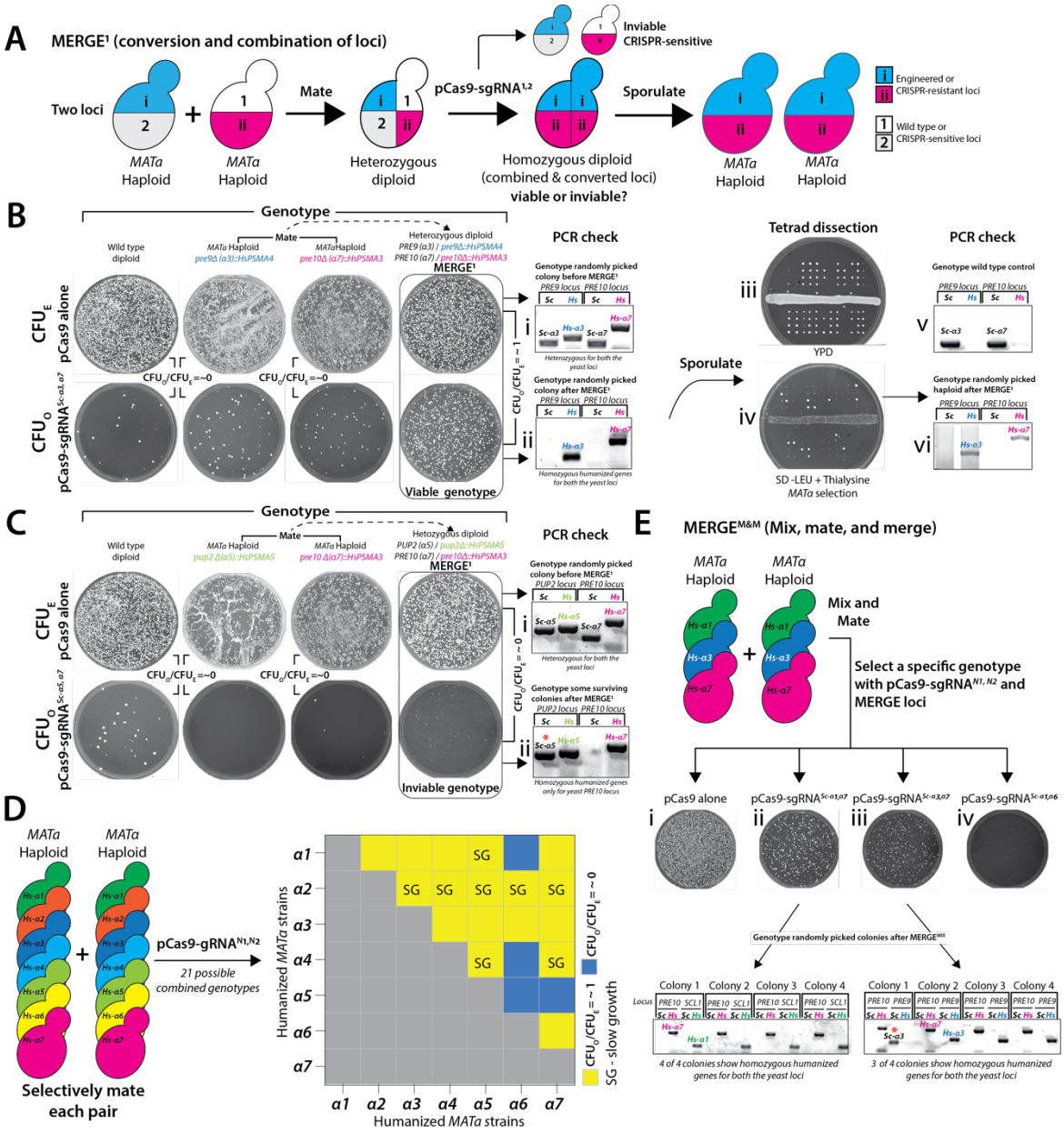


Figure 2.3. MERGE level 1 (MERGE[1]) efficiently converts and combines two heterozygous to homozygous loci.

(A) The schematic shows the mating of haploid yeast strains, each harboring two different alleles, enabling the combination as a heterozygote for each locus. Double-sgRNA CRISPR reagent targeted to both the alleles (CRISPR-sensitive loci, 1 & 2) initiates recombination at both loci using the homologous chromosome with CRISPR-resistant loci (i & ii) as a repair template, thus, enabling the simultaneous combination of 2 loci. $CFU_O/CFU_E = \sim 1$ suggests viable combined genotypes. Sporulation of the resulting homozygous diploid yields the desired combined genotype in both mating types. (B) Double-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\alpha3,\alpha7}$) targeted to two wild type yeast loci (proteasome $\alpha3$ and $\alpha7$ genes) quantifies the efficiency of MERGE[1]. The transformation of pCas9-sgRNA$^{Sc\alpha3,\alpha7}$ in the wild type, and single-humanized haploid Hs$\alpha3$ or Hs$\alpha7$ strains is lethal, with few surviving colonies ($CFU_O/CFU_E = \sim 0$). However, the transformation of pCas9-sgRNA$^{Sc\alpha3,\alpha7}$ in a heterozygous diploid humanized strain shows no lethality ($CFU_O/CFU_E = \sim 1$). PCR-based genotyping of surviving diploid strains after MERGE[1] shows conversion of both yeast to the humanized loci (ii) compared to before MERGE[1] (i). Sporulation generated the paired-humanized haploids in

both mating-types (shown here as Tetrads spotted on YPD (iii) and *MAT α* selection-iv). PCR check of haploid strains confirms the combination of humanized loci (vi). PCR of wild-type yeast loci is shown as a reference (v). (C)Alternatively, double-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\alpha5,\alpha7}$) targeting two wild type yeast loci (proteasome *α*5 and *α*7 genes) in a corresponding heterozygous diploid humanized strain showed lethality, suggesting incompatible combination (CFU_O/CFU_E = ~0). The transformation of pCas9-sgRNA$^{Sc\alpha5,\alpha7}$ in the wild type, and single-humanized Hs*α5* or Hs*α7* strains is lethal, with few surviving colonies, suggesting ON-target activity (CFU_O/CFU_E = ~0). PCR-based genotyping of surviving colonies after MERGE^1 shows the conversion of only one (*α5*) yeast to the humanized locus (ii) compared to before MERGE^1 showing heterozygous genotype at both loci (i). (D) MERGE^0 generated 7 humanized *α* proteasome strains in each mating-type. Mating each single-humanized strain allows a systematic test for every viable double-humanized genotype (21 combined genotypes). Transformation of double-sgRNA CRISPR reagent uniquely targeting yeast genes in each paired heterozygote facilitates the combination of two humanized loci as homozygotes (CFU_O/CFU_E = ~1, yellow) while also revealing genotypes that are not permitted (CFU_O/CFU_E = ~0, blue). (E) Mixing and mating singly humanized genotypes (Hs*α*1, Hs*α*3 and Hs*α*7) allows a random combination of two humanized alleles after double-sgRNA CRISPR selection referred to as mix, mate, and MERGE (MERGE$^{M\&M}$). Transformation of pCas9 alone serves as a control (i). Double-sgRNA CRISPR reagents, pCas9-sgRNA$^{Sc\alpha1,\alpha7}$ (ii) and pCas9-sgRNA$^{Sc\alpha3,\alpha7}$ (iii), specifically enriched the corresponding double-humanized genotype while also converting yeast to humanized loci. PCR-based genotyping of randomly picked colonies confirms *Hsα1α7*(4 of 4 colonies tested) and *Hsα3α7* (3 of 4 colonies tested). Comparatively, the transformation of double-sgRNA CRISPR reagent, pCas9-sgRNA$^{Sc\alpha1,\alpha6}$ (iv), shows no surviving colonies as the genotype does not exist in the mixture.

2.2.4 MERGE is scalable to combine multiple genetically engineered loci

The CRISPR-Cas9 system enables multiplexed editing by introducing multiple sgRNAs (Cong et al., 2013). To test scalability, we designed MERGE^{MX} (MERGE multiplex) to verify if more than two genetic loci can simultaneously convert to engineered loci by building the 4-gene carotenoid biosynthesis pathway from the carotenogenic yeast *Xanthophyllomyces dendrorhous* into baker's yeast (**Figure 2.4A**). The carotenoid pathway provides a colony color readout as a proxy for pathway engineering (**Figure 2.4A**). MERGE^0 generated haploid strains of opposite mating-types for each carotenoid transcription unit (Figure S2.5C), MERGE^1 provided the double-carotenoid genotypes (**Figure 2.4B**), and MERGE^{MX} enriched a complete homozygous carotenoid pathway genotype (**Figure 2.4C**). Furthermore, genotyping of randomly picked dark orange colonies confirmed the conversion and combination of the engineered loci (**Figure 2.4C**).

To test whether MERGE^{MX} can perform the combination of >2 humanized α proteasome genotypes, we mated a previously obtained *Hsα1α2α3* strain (MERGE^1) with a wild-type strain generating a heterozygous diploid for all three loci. The strain provided a platform to test MERGE^{MX} by

simultaneously combining 3 distinct humanized loci. The triple-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\text{-}\alpha1,\alpha2,\alpha3}$) converted all yeast loci to human versions (7 of 9 colonies tested) (**Figure 4D**). Next, using MERGE^{MX}, we tested if a triple humanized *Hsα4α6α7* genotype is viable. Transformation of triple-sgRNA, pCas9-sgRNA$^{Sc\text{-}\alpha4\alpha6,\alpha7}$ in a mated mix of haploid *Hsα4α7* (*MAT a*) and *Hsα6* (*MATα*) strains allowed simultaneous conversion of three wild-type yeast loci to humanized versions without using any diploid specific selection (**Figure 2.4E** & S2.11A). We found that if a triple-humanized intermediate genotype is viable, MERGE^{MX} can enrich and combine the specific genotype. To further test the ability of MERGE to scale to >10 loci, we designed a CRISPR reagent targeting a *GFP* expression cassette in a GM strain (Suzuki et al., 2011) (16 *GFP* loci, Figure S2.9A). The transformation of pCas9-sgRNA^{GFP} resulted in a few survivors that failed to show *GFP* expression, likely due to mutations at *GFP* loci due to NHEJ (Figure S2.9B & S2.9C), suggesting successful targeting of the majority of *GFP* cassettes.

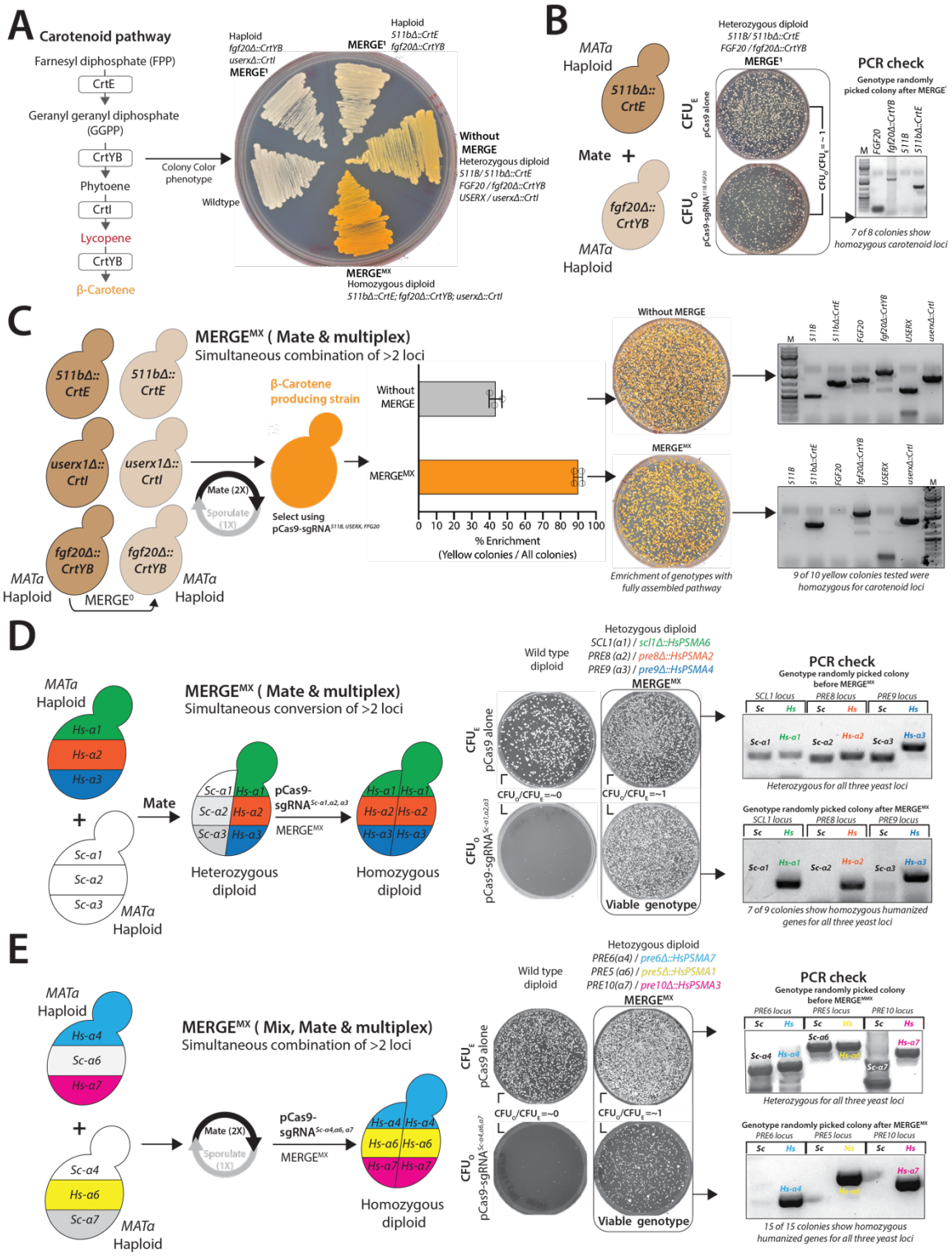


Figure 2.4. MERGE mate and multiplex (MERGE^{MX}) is scalable to combine multiple engineered loci.

(A) Schematic shows the carotenoid pathway genes and the metabolic intermediates leading to the color colony phenotype. A complete pathway (*CrtE*, *CrtYB* & *CrtI*) leads to an orange colony appearance, whereas the partial assembly (*CrtYB* & *CrtI*) provides an off-white colony phenotype. MERGE[1] provided single carotenoid engineered strains in both mating types. The homozygous diploid for a complete carotenoid pathway shows an intense orange colony phenotype than the heterozygous strain. (B)Schematic showing MERGE[1] combining two engineered genotypes at landing pad loci. Double-sgRNA CRISPR reagent (pCas9-sgRNA^{511B,FGF20}) targeting two landing pad loci (*511B* and *FGF20*) quantifies the efficiency of MERGE[1]. PCR- based genotyping of several colonies after MERGE[1] shows the conversion to the engineered carotenoid loci. (C) Single-carotenoid gene strains of both mating-types were mixed, mated (2X) and sporulated (1X). (D) Schematic shows mating haploid triple-humanized *Hsα1,α2,α3* strain with a wild-type yeast enables the combination of 3 humanized loci as heterozygotes.(E) MERGE^{MX} similarly combined >2 loci after mating a strain with 2 humanized loci (*Hsα4,α7*, *MATa*) with a strain carrying 1 humanized locus (*Hsα6*, *MATα*). The transformation of pCas9-sgRNA^{Scα4,α6,α7} in the wild type is lethal, suggesting ON-target activity ($CFU_O/CFU_E = \sim 0$). However, the transformation of pCas9-sgRNA^{Scα4,α6,α7} in a mated mix selected a diploid triple-humanized strain ($CFU_O/CFU_E = \sim 1$). PCR-based genotyping of several colonies after MERGE^{MX} shows conversion of all yeast to the humanized loci.

2.2.5 MERGE enables fitness-driven engineering of a near entire human *a* proteasome core in yeast

To explore the proficiency of MERGE for testing many combinations of engineered loci, we asked if an entire yeast heptameric α proteasome ring is humanizable. As an alternate strategy, we also tested the feasibility of sequential engineering using repetitive CRISPR selections and exogenous human gene repair templates (**Figure 2.5A**). The co-transformation of a triple-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\text{-}\alpha 1\ \alpha 2\alpha 3}$) targeting yeast *α1*, *α2* and *α3* genes and PCR fragments of human gene repair templates was successful in obtaining a triple-humanized strain (Figure S2.10A-i). A similar strategy generated triple-humanized *Hsα1α2α4* and *Hsα1α3α4* strains but failed to obtain a quadruple-humanized *Hsα4α5α6α7* (Figure S2.10A-ii). Thus, yeast genes are replaceable sequentially either alone or as small-scale simultaneous replacements. Using a triple-humanized genotype (*Hsα1α2α3*), the humanization of yeast *α4* was successful. Next, we explored *Hsα5*, *Hsα6* and *Hsα7* humanizations in parallel; however, we obtained only one quadruple humanized *Hsα1α2α3α4α7* genotype (**Figure 2.5A**, Figure S2.10B). The functional replacement of yeast *α5* or *α6* was unsuccessful despite repeated attempts. The plasmid-borne expression Hsα6 in a

quintuple-humanized strain (*Hsα1α2α3α4α7*) resulted in a toxic phenotype (no growth), suggesting that further humanizations are incompatible (**Figure 2.5A**). Overall, while the sequential strategy was partly successful, it failed to reveal if the inability to humanize yeast *α* core entirely was due to incompatible genotypes or inefficient genome editing, especially as hybrid human-yeast genotypes show growth defects and reduced transformation efficiencies (Figure S2.10C). However, MERGE^{MX} provided a clear readout of incompatible humanized genotypes, readily generating many combinations of humanized genotypes (**Figure 2.5B**). We first explored several triple-humanized genotypes, obtaining *Hsα1α2α3* & *Hsα4α6α7* strains (**Figures 2.4D & 2.4E**) while *Hsα4α5α6* and *Hsα4α5α7* genotypes are inviable (**Figures 2.5B**, S2.11B & S2.11C). By mating yeast strains with distinct humanized genotypes, we next explored many higher-order (>3) combinations, obtaining quadruple-humanized *Hsα1α2α3α4* and *Hsα1α2α3α7* genotypes. In comparison, the quadruple-humanized *Hsα4α5α6α7* genotype is not feasible (**Figure 2.5B**, Figure S2.12A, S2.12B, S2.12C & S2.12D). Subsequent MERGE strategies generated viable quintuple humanized genotypes, *Hsα1α2α3α6α7* and *Hsα1α2α3α4α6,* whereas *Hsα1α2α3α4α5* genotype is inviable (Figures 2.5B, S2.12E, S2.12F, & S2.12G). The following MERGE^{MX} assay yielded a viable sextuple humanized *Hsα1α2α3α4α6α7* genotype with a delayed growth phenotype (Figure S2.11H). To conclusively verify if the entire yeast *α* proteasome core is humanizable, we mated partially humanized *Hsα1α2α3α6α7* and *Hsα4α5* strains, allowing the combination of all *a* proteasome genes as heterozygous human-yeast genotypes (**Figure 2.13A**). MERGE^{MX} converted yeast *α1*, *α2*, and *α3* loci to homozygous human alleles. However, the subsequent conversion of the remaining four yeast *α4α5α6α7 t*o human loci using pCas9-sgRNA$^{Sc\alpha4,\alpha5,\alpha6,\alpha7}$ did not yield viable colonies suggesting a fully human *α* proteasome core is incompatible (Figure S2.13B). Overall, MERGE successfully tested many combined humanized genotypes. This gradual

progression from yeast to humanized α proteasome core rescued the viability of specific incompatible double-humanized genotypes, suggesting that these subunits are co-humanizable when neighboring interactions are restored (**Figures 2.3D** & S2.14). The data reveals that the proteasome subunits *α5* and *α6* are not co-humanizable in yeast (**Figure 2.5**). The sporulation failure observed in genomically-replaced strains can lead to dead ends while performing MERGE (associated with all humanized *Hsα7* genotypes, Figures S2.11A, S2.12A& S2.12F). We propose two solutions: One, by allowing the strains with heterozygous engineered loci to sporulate without MERGE, followed by CRISPR plasmid selection to enrich combined haploid genotypes (Figure S2.15A). Alternatively, a sequential strategy can engineer viable genotypes in a haploid strain (Figures S2.12F & S2.15B). Furthermore, using CRISPR reagents to generate multiple DSBs at several yeast loci could potentially have OFF-target effects. Therefore, we performed whole-genome sequencing of singly humanized *(Hs-α1)* and quintuple-humanized strains (*Hsα1α2α3α4α7*), ruling out OFF-target DSBs and mutations (Figure S2.16).

Proteasome biogenesis is a highly regulated process aided by several assembly chaperones (Budenholzer et al., 2017; Hirano et al, 2006; Le Tallec et al., 2007). In the case of α proteasome core, particularly α5 and α6, subunits interact with assembly chaperones, enabling ordered assembly (Schnell et al., 2021; Morrris et al.,, 2021; Stadtmueller et al., 2012). We analyzed the expression of human proteasome subunits by mass spectrometry (MS) and found proteome-wide changes with significant over-expression of many yeast proteasome subunits, including several assembly chaperones in the humanized proteasome strains (Figure S2.17A & Table S2.2). The β core assembly immediately follows the α core, and the incompatible interface may now require human β subunits in yeast (Budenholzer et al., 2017; Murata et al., 2009). The heterologous expression of human assembly chaperones or human β subunits in a humanized α core strain may

permit the synthesis of a fully human catalytic core particle in yeast. We performed the co-expression of human constitutive- and immuno-βs followed by MS analysis in the singly humanized Hsα1, quintuple-humanized *Hsα1α2α3α4α7* and sextuple-humanized *Hsα1α2α3α4α7β3* (also harboring human β3 in place of the yeast ortholog) strains showing stable expression of human β subunits only in partially humanized yeast strains (Figure S2.17A). Together with the phenotypic rescue of the humanized yeast strains by human β subunits, these data suggest the assembly of a human 20S proteasome core in yeast (Figure S2.17B & S2.17C).

Given the complex synthesis and architecture of the proteasome (Tomko & Hochstrasser, 2013), it is challenging to know if there are a limited number of 'paths' to engineer a fully humanized 20S core particle due to a rapidly accumulating number of assays to perform (as in Figure S2.1). However, using MERGE and the co-expression of critical human proteasome subunits (as in Figure S2.17), an entirely humanized catalytically active yeast 20S proteasome core is possible.

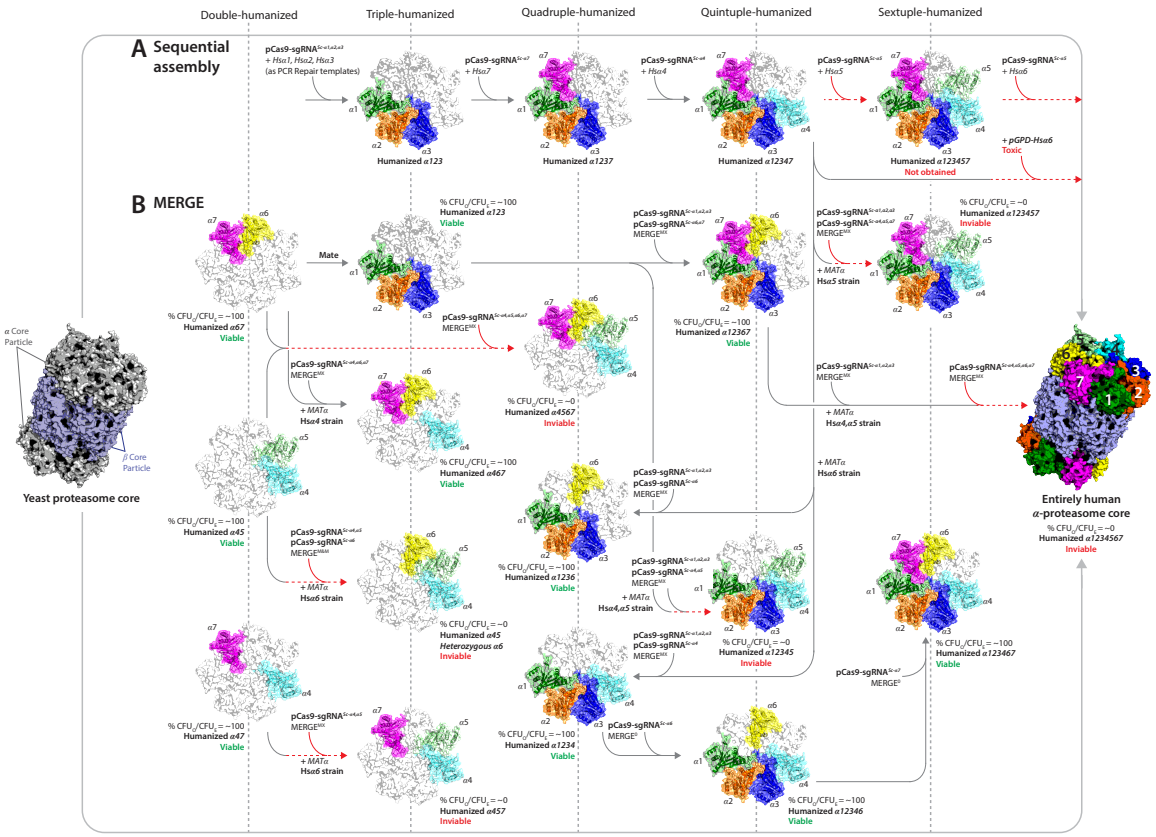


Figure 2.5. MERGE reveals a fitness-driven path to the humanization of most of the *α* proteasome core in yeast.

(A) The schematic shows the transition of heptameric yeast *α* proteasome to humanized *α* proteasome core. (A) A sequential strategy used a triple-humanized *Hsα1,α2,α3* strain as a background to progressively humanize the rest of yeast *α* core genes by co-transforming a CRISPR reagent and a human gene repair template. Choices to sequentially humanize the *α* proteasome core were made depending on the success of the prior effort. The strategy permitted the humanization of yeast *α7* followed by *α4.* Whereas several attempts to humanize yeast *α5* and *α*6 failed. The expression of human Hsα6 in a quintuple-humanized Hsα1,α2α3,α4,α7 strain resulted in a lethal phenotype. (B) MERGE provided a clear readout of the fitness of combined genotypes while revealing incompatible combinations of the humanized yeast α proteasome core. Mating distinct humanized yeast combinations (show as connecting lines) followed by MERGE^{MX} tested various triple-, quadruple-, quintuple- and sextuple-humanized *α*-proteasome yeast strains. MERGE generated viable triple-humanized combinations *Hsα4,α5,α6* and *Hsα4,α5,α7* (indicated as dashed lines in red), suggesting incompatible combinations. Similarly, while MERGE^{1}facilitated the generation of *Hsα4,α5* and *Hsα6,α7* genotypes, the subsequent MERGE^{MX} to generate quadrupled *Hsα4,α5,α6,α7* failed. Using MERGE^{MX} followed by MERGE^{0} identified several quadruples, quintuples and one sextuple-humanized genotype (*Hsα1,α2,α3,α4,α6,α7),* revealing a fitness-driven path to the humanization of near entire yeast α-proteasome core. Whereas identifying humanized combinations (>2), comprising the Hsα5 subunit, as incompatible genotypes. Mating of *Hsα1,α2,α3,α6,α7* and *Hsα4,α5* strains generated a heterozygous human-yeast diploid for all 7 α proteasome core genes. Using multiplex-sgRNA CRISPR reagents permitted verifying if the entire human α-proteasome core is feasible. A triple-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\text{-}\alpha1,\alpha2,\alpha3}$) homozygosed 3 of 7 humanized loci. However, a subsequent MERGE^{MX}, using a quadruple-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\text{-}\alpha4,\alpha5,\alpha6,\alpha7}$), failed to obtain a viable genotype. Proteasome core structures were generated using Pymol and PDB- 1RYP. Colored structures show humanized α yeast subunits.

2.3 Discussion

By integrating marker-less CRISPR-Cas9-based selection with cycles of mating and sporulation, we developed a new method (MERGE) that enables efficient combination of multiple distinct loci. The advantages of using CRISPR selection are numerous, with no requirement of markers for any number of genotypes. Given the efficiency of MERGE, and because the method entails the creation of viable heterozygous intermediate strains, it permits the quantitative identification of all possible combined genotypes. MERGE can probe higher-order genetic interactions (>3 loci) limited by the availability of selection markers in budding yeast (Kuzmin et al., 2021), and given the broad conservation DNA repair machinery, MERGE can simplify genetic assays in model systems, such as *Candida albicans* and *fission yeast.*

In conclusion, MERGE offers highly scalable multi-locus genome engineering in diploid yeast cells by using a high-efficiency CRISPR-based gene-drive-like strategy to overcome the independent assortment of unlinked loci. MERGE has the potential to allow systematic functional genomic analysis in other systems lacking sophisticated tools and drive synthetic and systems biology research from engineering heterologous systems to performing multi-site and genome wide combinatorial editing in yeast, as we demonstrated by engineering 89 independent sites along Chromosome 1, a complete four gene carotenoid biosynthesis pathway, 16 *GFP* insertions within the same strain, and a majority of the α proteasome genes. By engineering 6 of 7 human α proteasome core genes in yeast, our work also demonstrates the remarkable degree of functional conservation in the proteasome complex despite over a billion years of evolutionary divergence, extending from a single gene to nearly an entire module. The data confirm our previous

observations that humanization seems to be driven by modules of physically or functionally interacting proteins being similarly replaceable (Kachroo et al., 2015). Further characterization of the incompatibilities should reveal novel orthogonal functions or interactions in diverged species. However, pursuing a combinatorial strategy with MERGE along with a sequential strategy in parallel allows one to inform the other about simultaneous replacements that are likely to work. Humanizing all or multiple members of a protein complex will allow a novel approach to learning human biology, including complex assembly, biogenesis, and variant effects on function, investigations of their contributions to disease, and the possibility of seeking therapies for these diseases in the simplified context of a yeast cell.

2.4 Materials & methods

2.4.1. Targeting sgRNA design

CRISPR-Cas9 targeting sequences consist of a 20 bp recognition sequence preceding an 'NGG' sequence motif (PAM). To design our targeting sequences, we used the built-in guide RNA design tool in version 11 of the Geneious software (Kearse et al., 2012), using a recent version of the yeast genome (available from http://www.yeastgenome.org/strain/S288C/overview) to screen for possible off-target sequences (Akhmetov et al., 2018; Kachroo et al., 2017). We chose two high scoring guide sequences to target each gene, requiring that they be near the 5' end of the gene so any NHEJ repair would likely result in an early frameshift. The guides were ordered as complementary oligos with overhangs according to the sgRNA template described in the Yeast Tool Kit (Lee et al., 2015) (Table S2.1).

2.4.2. Plasmid construction using Yeast Tool Kit (YTK)

Entry vectors for each guide RNA sequence were constructed by the BsmbI-mediated Golden Gate (GG) assembly (Thermo) into plasmid pYTK050 from the Yeast ToolKit (Lee et al., 2015). The targeting sequences were ordered (IDT) as two single-stranded DNA oligos with a complementary region and unique overhangs according to the 'sgRNA' template described in Lee, et al., (Lee et al., 2015). Briefly, the two oligos for each guide were annealed by slow cooling from 95°C to 4°C (1-5 hrs) in a thermal cycler, and 1µl of the annealed product was added to the entry vector GG reaction. Entry Vectors were sequence-verified using custom primers. 'Transcriptional unit' (TU) vectors (pYTK095) were constructed by a BsaI-mediated GG assembly (Thermo). The entry vector for a particular guide RNA was combined with a left connector part plasmid (ConLX) and right connector part plasmid (ConRX) into the AmpR-ColE1 (pYTK095) backbone plasmid (Akhmetov et al., 2018; Kachroo et al., 2017). To create single sgRNA TU vectors, we assembled the sgRNA entry vector with left connector 1 (ConL1, pYTK003) and right connector E (ConRE, pYTK072). To create TU vectors for the multi-sgRNA knockout, an entry vector was combined with the appropriate connectors to enable proper subsequence knockout vector assembly (i.e. a triple-sgRNA vector would have one TU vector with ConL1 and R2, one with ConL2 and R2, and the third with ConL3 and RE). 'Knockout' (KO) vectors were constructed by BsmbI-mediated GG assembly (Thermo). The appropriate TU vectors were assembled into the *CEN6-URA*3 backbone and the Cas9-TU1 vector to create a KO vector. Each GG reaction was performed in a 10µl volume, with approximately 20fmol of each starting DNA molecule, along with 0.5µl each of the appropriate restriction enzyme, 0.5µl of T7 DNA ligase (NEB), and 1µl of 10X T4 ligase buffer (NEB). A thermocycler was used to cycle between 16°C and 37°C each for (5 minutes). Three µl of the reaction volume was transformed into competent DH5α *E. coli* and plated on appropriate selective media for the backbone (*i.e.,* Chloramphenicol for entry vectors, ampicillin for TU vectors, and

kanamycin for KO vectors). Each backbone plasmid is a '*GFP*-dropout' vector, so correct clones were selected by screening for non-fluorescent colonies when viewed by blue light and verified by sequencing.

2.4.3 Cas9-TU1

The Cas9-TU1 vector was constructed by *BsaI* GG assembly of YTK parts pYTK002 (ConLS), 011 (*PGK1* promoter), 036 (Cas9 coding sequence), 055 (*ENO2* terminator, 067 (ConR1), and 095 (*AmpR*-ColE1 *GFP* dropout backbone).

2.4.4. *CEN6-URA3* expression vector

The *CEN6-URA GFP* dropout backbone vector was constructed by BsaI GG assembly of YTK parts pYTK008 (ConLS'), pYTK047 (*GFP* dropout), pYTK073 (ConLE'), pYTK074 (*URA3*), pYTK081 (*CEN6/ARS4*), pYTK084 (*KanR*-ColE1 RFP dropout vector).

2.4.5. An alternate strategy to directly clone annealed sgRNAs in a yeast expression vector

To perform faster direct sgRNA-Cas9 plasmid construction without requiring three independent cloning strategies (as described in section 2), we constructed pCas9-GFPdo to directly clone the annealed sgRNA primers into yeast expression vectors. The sgRNA expression unit from the pYTK050 plasmid was amplified via PCR using primers with BsaI sites that generated overhangs similar to ConL1 (forward primer) and ConRE (reverse primer) to clone in *CEN6-URA* or CEN6-*kanMX* yeast expression vectors. Finally, using the GG reaction protocol, the sgRNA TU PCR was assembled into the *CEN6-URA3* or CEN6-*kanMX* backbone with the Cas9-TU1 vector to create a direct knockout (pCas9-*GFP*do) vector.

2.4.6. An alternate strategy to directly clone multiple sgRNAs TUs in a yeast expression vector for multiplexing in MERGE^{MX}

The YTK kit allows cloning multiple sgRNA expression cassettes in a yeast expression vector. However, the strategy involves three cloning steps. To build >1 sgRNAs TUs directly into the yeast expression vector without generating intermediate connector vectors, we designed the pCas9-*GFP*MX (*CEN6-URA3*) vector. Primers were designed to PCR amplify the *GFP* expression cassette (for *E. coli* expression, pYTK047) with GG enzyme sites to clone *GFP* (BsaI for cloning the *GFP* insert, the sites are lost after GG cloning). The BsmBI sites, with overhangs similar to ConLS and ConRE, designed in the primers, allow the PCR-based sgRNA TU cloning (obtained by ligating annealed sgRNA primers in pYTK050) in *CEN6-URA3* yeast expression vector (Table S1). The primers used to amplify the sgRNA expression units now harbor connector overhangs to clone >1 sgRNA expression units in tandem, eliminating the need to individually make each connector clone.

2.4.7. Repair design and construction

Repair DNA was designed to be a linear DNA molecule that contained the human coding sequence, from the start codon to stop, with at least 100bp of flanking homology to the yeast genome immediately upstream and downstream of the native yeast start and stop codons (Figure S3B). They were constructed by PCR using human ORFeome (Lamesch et al., 2007) or MGC (MGC Project Team et al., 2009) clones as templates, using primers with long extensions providing the homology sequence. Repair template PCRs were performed with Accuprime Pfx (Thermo) as multiple 100µl reactions according to the manufacturer's protocol, combined and purified using the Zymo DNA

Clean or Qiagen PCR Cleanup and Concentrator -25 kit. Approximately 5µg of repair DNA was included in the transformation mix during sequential replacement transformations.

2.4.8. Humanization procedure

Yeast transformations were performed using the Frozen-EZ Yeast Transformation 846 II Kit from Zymo Research (Cedarlane). In short, cells were grown to mid-log phase, washed with kit Solution 1, resuspended in kit Solution 2 and either used directly to transform or frozen at -80 C for future use. In general, we transformed approximately 1µg of the KO plasmid along with 5µg of repair DNA, which yielded anywhere from single to dozens of colonies, depending on transformation efficiency and the number of simultaneous molecules necessary. In general, nearly all screened clones were successful. For initial single replacement in the wild-type strains, we tested two sgRNA sequences for each gene. These were tested in parallel, together with a no repair control transformation to assay for sgRNA effectiveness. Successful sgRNAs show zero or very few colonies in the no-repair control transformation (CFU_O/CFU_E =~0, Figure 2.1A & S2.4A).

2.4.9. Engineering carotenoid genes at the landing pad loci

Single-sgRNA CRISPR reagents targeting three landing pad loci (*511B, USERX, FGF20*) (Bourgeois et al., 2018) were generated as described in section 1. The repair templates were generated as previously described in section 7. Specifically, carotenoid gene transcription units were amplified from plasmids harboring the transcription units using the primers with 80bp homology to the landing pad loci (Mitchell et al., 2015) (Table S2.1). Each landing pad locus was edited in the SGA strain (*MATa*) to harbor a carotenoid gene transcription unit (*511B* for *CrtE*, *USERX* for *CrtI,* and *FGF20* for *CrtYB*) driven by strong constitutive yeast promoters. The

transformation was carried out using the Zymogen frozen EZ yeast transformation kit (Cedarlane), and colonies were confirmed by PCR using primers specific to the landing pad or the carotenoid gene loci as described below (Figure S2.4B & Table S2.1).

2.4.10. Clone verification

Clones were initially screened by colony PCR using a rapid DNA isolation method and colony PCR (Lõoke et al., 2011). Forward primers for PCR screening were designed such that the upstream primer would bind in the yeast genome approximately 150bp - 500bp upstream of the yeast ORF or the landing pad loci, to ensure it was outside of the homology region used for repair. the human or carotenoid gene sequence. To facilitate multiplex PCR screening, each pair amplifies different-sized bands for the yeast and human or carotenoid genes. Following plasmid loss, clones were further verified by directly sequencing a PCR product using Sanger sequencing.

2.4.11. Plasmid loss to perform sequential editing of loci

Successful clones were subjected to a plasmid loss procedure to alleviate any stress incurred by the constitutive expression of the Cas9 protein and allow further knockout plasmids to be transformed. To avoid any potential stress-induced defects, we rarely used 5-FOA counter selection to force loss of the *URA3* plasmid, instead relying on spontaneous plasmid loss. One two methods were used to identify spontaneous plasmid loss. Clones were grown overnight in YPD and then again either spread onto YPD (100µl of 1:1000 dilution) and replica plated on SC -Ura and YPD or patch plated (typically 6-12 colonies) to both minimal media lacking uracil and YPD. This procedure resulted in around 10-60% of colonies losing the plasmid (estimated). For a plasmid harboring *URA3* selection, the strains that lost the plasmid were identified using 5- FOA selection.

2.4.12. Growth curves

Cells were diluted to approximately 0.01-0.02 OD^{660} (~2-5 $x10^5$ cells) in 150µl YPD across 3-4 replicates. They were grown and read in a BioTek Synergy H1 or Sunrise Tecan microplate reader, with continuous double orbital shaking at 30℃ or 37℃. Reads were taken while stationary every 10-20 minutes, and experiments were run for at least 24-36 hrs. For spot assays, strains were grown overnight in YPD at 30℃ and spotted with serial dilutions on YPD agar. The plates were incubated for 2-3 days at 23℃, 30℃ and 37℃.

2.4.13. CRISPR-based selection to identify and enrich unique genotypes

To test the efficiency of CRISPR selection, a subset of engineered strains (Haploid *MATa*/α) were mixed as a pool. Each strain was cultured overnight in YPD individually and back diluted into a pool the next day at equal OD and grown to mid-log phase. The strains used in the mix were all the same mating types to avoid random mating events during the experiment. Competent cells were prepared and transformed with various pCas9-$sgRNA^{locus}$ plasmids. We performed the experiments as biological and technical replicates using different genotypes in the mix. The same competent cell mix was either transformed with pCas9 alone or with various pCas9-$sgRNA^{WT}$ locus vectors, each harboring a different yeast-specific sgRNA expression unit. Competent cells were generated using Zymo Research Frozen EZ kit transformation protocol. Each transformation was plated on SD-URA. pCas9 alone served as a positive control allowing the growth of all genotypes in the mixture. pCas9-$sgRNA^{ADE2}$ was used to visually observe and quantify the efficiency of the selection. The plasmid will eliminate all the other genotypes except the *ade2Δ::kanMX* (red colonies due to the disruption of the *ADE2* gene). pCas9-$sgRNA^{RPT5}$ was used as a negative control as the yeast *Rpt5*

gene is present in all the strains such that yeast cells should survive this transformation. Colonies from each Petri plate were picked randomly and genotyped by locus-specific PCR to confirm the CRISPR reagent exclusively selected a unique genotype as described in section 10. (**Figure 2.1C**).

2.4.14. Mating and MERGE^0 assays

The mating assays were carried out in the following ways: The engineered strains with single gene modification in BY4741 (Haploid *MATa*) backgrounds were mated with SGA-haploid strains harboring haploid specific markers (SGA *MATa*; *can1Δ::STE2pr_Sp_his5his3Δ1leu2Δ0ura3Δ0*; select on SD-His+Canavanine & SGA *MATα*; *lyp1Δ::STE3pr_LEU2 his3Δ1leu2Δ0ura3Δ0*; select on SD-Leu+Thialysine). Each parental strain was transformed with two different empty vector plasmids with distinct selectable markers *URA3* and *kanMX,* and mated diploids were selected on SD-Ura+G418. The haploid strains were grown overnight in 5ml selection media (SC -Ura or YPD +G418). The strains were mixed in a rich medium and incubated at 30°C for 3-4 hrs. with shaking. 500µl of the mated mixture was washed with distilled water, and 10µl of the mix was plated at different dilutions on solid agar media using SD -*URA*+G418 selection and incubated for 2-3 days at 30°C. Plasmids were cured using a previously described strategy in section 11. The heterozygous colonies were confirmed by PCR genotyping using primers specific to the wild-type and engineered loci as described in section 10. The confirmed heterozygous strains for all proteasome and carotenoid strains were transformed with CRISPR plasmids targeting wild-type loci and converting the parent heterozygous strain into homozygous diploid strain. Colonies were picked randomly from the plate and genotyped for humanized or carotenoid loci.

Alternatively, instead of an empty vector, the engineered humanized or carotenoid strains were transformed with a CRISPR reagent for which the corresponding strain harbors a CRISPR-resistant

locus. The subsequent mating with an opposite mating-type strain with an empty vector (pCas9 alone or a backbone plasmid) results in the loss of heterozygosity and conversion of single loci ($MERGE^0$). In each scenario, the number of colonies obtained on plates transformed with CRISPR reagent (CFU_O) were compared with empty vector transformations (CFU_E) to calculate the efficiency of $MERGE^0$. Each experiment was performed at least 3 times.

2.4.15. Sporulation of diploid and selection of haploid yeast strains

Diploid strains were sporulated in a media containing 0.1% (w/v) potassium acetate and 0.005% (w/v) zinc acetate for 4-7 days. For tetrad dissection, spores were spun down at 5000rpm for 5 minutes, resuspended in 200µl of 20mg/ml Zymolyase and incubated at 37^{o}C for 25-30 minutes. The mix was then incubated at -20^{o}C to stop the reaction. Cells were thawed on ice, and 20ul of the mix was plated into YPD for Tetrad dissection. Tetrad dissection was performed using Spore play (Singer Instruments for every engineered proteasome spore on YPD and then replicated on SGA selection. For carotenoid and humanized loci in SGA background strain, the sporulation mix was directly plated on SGA selection. Locus-specific PCR verified the haploid strains for the presence or absence of engineered loci.

2.4.16. Estimating the efficiency of $MERGE^0$ using yeast *ADE2* locus as a readout

The yeast *ADE2* locus was used to estimate $MERGE^0$ using a CRISPR-resistant allele *ade2Δ::kanMX* on the homologous chromosome as a repair template. The sgRNAs were designed using Geneious version 11 as described in section 1 (Table S2.1). The pCas9-sgRNA*ADE2* vector showed ON-target activity causing lethality in wild-type haploid or diploid strains. Occasional colonies show a red colony phenotype suggesting mutations via NHEJ at the *ADE2* locus. To

estimate MERGE^0 efficiency, the pCas9-sgRNA^{ADE2} was transformed in a diploid hetKO *ADE2/ade2Δ::kanMX* obtained from the yeast "Magic Marker" hetKO collection(Pan et al. , 2004). To further verify that the red colonies observed after MERGE^0 are not due to the mutations in the *ADE2* locus (NHEJ) and instead due to conversion to *ade2*Δ::*kanMX* locus (HDR), the heterozygous diploid (*ADE2/ade2Δ::kanMX*) strain transformed with a control (pCas9 alone) and pCas9-sgRNA^{ADE2} plasmid were sporulated followed by tetrad dissection (Figure S2.5A). Haploid spores were selected on YPD or YPD + G418 (200μg/ml). Spores harboring a control plasmid showed the expected 2 : 2 (red : white phenotype) and 2 : 0 (G418 resistant, G418 sensitive) phenotype compared to the pCas9-sgRNA^{ADE2} transformed cells that show 4 : 0 (red, white) colonies and 4 : 0 (G418 resistant, G418 sensitive) colonies phenotype, respectively, suggesting the conversion to *ade2Δ::kanMX* rather than the mutation of *ADE2* locus. In contrast, the co-transformation of pCas9-sgRNA^{ADE2} and oligo as a repair template harboring (100bp homology, 5X molar excess than the plasmid) to the 5' and 3' UTRs of *ADE2* locus in haploid wild-type yeast cells, shows resulted in significantly fewer survivors (Figure 2C; % CFU_O/CFU_E = 21.6+/-SD; N=4). However, this method is still far less efficient than MERGE^0 (**Figure 2.2C**; ~100% vs 21.6%).

2.4.17. Using MERGE^0 to perform one-step gene essentiality assays in yeast

HetKO diploid strains for 7 α proteasome core genes were obtained from the yeast "MagicMarker" hetKO collection (Pan et al., 2004). The strains were transformed with either a single sgRNA CRISPR reagent targeting the corresponding yeast α proteasome genes or the empty vector control and selected on SD -URA+G418. CRISPR reagent transformed plates caused lethality in 6 of 7

α proteasome hetKO strains, except in the case of non-essential *a*3. A similar assay in the hetKO *ADE2/ade2Δ::kanMX* strain showed viable homozygous null cells for a non-essential *ADE2* locus.

2.4.18. Testing MERGE^0 efficiency for all yeast loci on Chromosome I

All the strains harboring heterozygous knockout diploid loci from the yeast "Magic Marker" collection on chromosome I were arrayed in a 96-well format (Pan et al., 2004) (Figure S6A). A2 is the locus close to the left telomere, F1 near the centromere, and G10 close to the right telomere (**Figures 2.2D** & S2.6A). The sgRNA targeting the *kanMX* cassette was designed using the strategy mentioned earlier (Table S1). Since the bacterial selection for the previously described (section 4) yeast shuttle vector harbors a KanR cassette identical to the *kanMX* cassette in the case of hetKO strains, we designed a new expression vector with AmpR selection. The plasmid showed a lethal phenotype in the strains harboring the *kanMX* allele (haploid or homozygous diploid) in the genome (**Figure 2.2D**). Approximately 1µg of pCas-sgRNA^{kanMX} was transformed for each transformation. First, we performed a pilot assay using 5 hetKO strains with genes located at various regions on chromosome I (Figure S6B). These strains showed no lethality (CFU_O/CFU_E of ~1). Each strain was inoculated in 800µl of *YPD* + His (50 mg/L) + G418 (200 µg/ml) overnight in 96- well format. The following day, cultures were back diluted and grown to the mid-log phase. Competent cells were generated in a 96-well format using the Gietz yeast transformation protocol (Gietz, 2007). Each transformation mix was transformed with either 1µg of pCas9 alone or 1µg of the pCas9-sgRNA^{kanMX} plasmid. An equal amount of the transformed cells were plated onto SD URA and SD-URA + G418. All of the yeast loci (except for A10- *CNE1*) on chromosome I show MERGE^0 mediated conversion of *kanMX* to wild-type yeast locus irrespective of the position of the gene

along the chromosome (CFU_O/CFU_E =~1) and simultaneously lost the *kanMX* cassette (Figure S2.6A).

2.4.19. Performing MERGE^1 and MERGE^{MX} at 2 or more loci using humanized proteasome and carotenoid strains

CRISPR plasmids expressing 2 or more sgRNA cassettes in tandem were designed and constructed as described in sections 1 - 6. Each single-humanized *a* proteasome (BY4741, *MATa*) was mated with SGA strain (*MATα*) as described in section 14. Heterozygous strains from each mating mix were confirmed by locus-specific PCR genotyping as in section 10. MERGE^0 enabled the conversion to humanized loci. The next round of mating in SGA strain (*MATa*) followed by MERGE^0 generated humanized strains in both mating-type SGA backgrounds. Carotenoid strains were directly generated using SGA compatible strains described in section 9, followed by MERGE^0 to obtain strains in both mating-types. Each double-sgRNA CRISPR reagent was lethal in singly engineered strains (Figure S2.7B). Mating between each distinct single-engineered genotype combines the loci as heterozygotes. Transformation of 0.5-1µg of double-sgRNA CRISPR reagent in heterozygous diploid strains either showed no lethality (viable genotype, CFU_O/CFU_E =~1) or a lethal phenotype (inviable genotype, CFU_O/CFU_E =~0) (**Figures 2.3B**, 2.3C &S2.7C). Homozygosity was confirmed by randomly picking colonies after MERGE^1 and performing a locus-specific PCR confirming the presence or absence of yeast or engineered genotype. For MERGE^{MX} to simultaneously target >2 loci, the CRISPR reagents were generated as described in sections 2-4 and 6. The haploid strains harboring combined engineering loci were mated with either a wild-type or strains carrying a different set of engineered loci. The diploids were transformed with either an empty vector or a multiplex-sgRNA CRISPR reagent. The genotypes were confirmed

by locus-specific PCR as described in section 10. Human β3 was genomically inserted using CRISPR-Cas9-sgRNA strategy, targeting the yeast Sc*PUP3* locus, in a quintuple-humanized *Hsα1,α2,α3,α4,α7* strain and verified using locus specific PCR and WGS.

2.4.20. Performing MERGE$^{M\&M}$ to combine 2 loci from the mated mix of single-humanized proteasome genes.

Single-humanized *α* proteasome strains *Hsα1*, *Hsα3*, and *Hsα7* of both mating-types were allowed to mate randomly in a mix. Each strain was inoculated at 0.3 OD^{600} in 2 ml YPD overnight, then added to a mix at equal OD^{600} and incubated on a shaking incubator for 4- 6 hrs. at 30°C. Five hundred ul of the mix was further cultured overnight at starting 0.3 OD, followed by competent cell preparation. The mixture was transformed with either an empty vector or double-sgRNA CRISPR reagent. Several colonies were observed post MERGE$^{M\&M}$ for *Hsα1α7*, and *Hsα3α7* paired genotypes. PCR-based genotyping confirmed the combination as a homozygous double-humanized diploid.

2.4.21. Testing the potential of CRISPR reagent to target multiple loci using the Green Monster strain.

To address the scalability of targeting multiple yeast loci, we used the Green Monster (GM) strain (Suzuki et al., 2011). The sgRNA target sequence was designed in *GFP* (S65T) variant using Geneious version 11 as described in section 1. pCas9-sgRNA^{GFP} was constructed using the method of direct cloning strategy as described in section 5 using a backbone with *AmpR* (*E. coli*) and *kanMX* (yeast) selection. The *MATα* mating-type Green Monster was used *MATα lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::GMToolkit-α [CMVpr-rtTA NatMX4 STE3pr-LEU2].* The OFF-*target*

activity of pCas9-sgRNA^{GFP} was tested in a haploid WT strain with no *GFP* gene showing a $CFUo/CFU_E$ =~ 1. The ON-target activity was measured transformation in natively tagged C terminal fusion of *GFP* to *BRO*1 gene (*BRO1-GFP*) obtained from the yeast *GFP* collection (Huh et al. , 2003). To estimate the loss-of-function mutations at the *GFP* locus, cells from an empty vector and pCas9-sgRNA*GFP* transformed GM yeast cells were subjected to flow cytometry (BD Accuri™ C6 Plus Flow Cytometer). Cells were grown in 96-well format in *YPD* overnight and back diluted in SC media with 10µg/ml of doxycycline for ~48 hrs. The culture was diluted 1 in 10 in water and passed through the flow cytometer. Using BY4741 yeast with no *GFP* as a control, the cells showed mostly background fluorescence as expected. GM cells transformed with an empty vector show high fluorescence as this strain harbors 16 *GFP* cassettes serving as a positive control. Finally, random colonies were picked and pooled from haploid GM cells transformed with pCas sgRNA^{GFP}, and the mixture passed through the flow cytometer. The cells show little *GFP* expression similar to the wild type, suggesting ON-target NHEJ-mediated mutations at nearly all *GFP* loci (Figures S2.8).

2.4.22. Performing MERGE^{MX} using mating and sporulation cycles to assemble an entire Carotenoid pathway in yeast.

SGA strains of both mating-types harboring single-carotenoid transcription units at the landing pad loci were generated in sections 2.4.9 and 2.4.14. The strains were cultured independently overnight, followed by mixing at a similar 0.3 OD^{600} and cultured on a shaking incubator for 5-6 hrs. A hole was made on the lid of the Eppendorf tube to allow aeration, and the mixture was incubated at 30°C for 12 -16 hrs. The formation of diploids was confirmed by light microscopy. Next, 500µl of culture was centrifuged at 3500 rpm for 5 minutes. The supernatant was removed and resuspended in 2 -

5ml of sporulation media, followed by incubation on a rotating shaker for 5-8 days at room temperature. Next, 500µl of sporulation mix was centrifuged at 3500 rpm for 5 min and treated with Zymolyase. After the first sexual cycle, the cells were centrifuged, washed with water, and resuspended in rich media YPD. The cycle was repeated to generate a second set of diploids. The mixture was incubated for 4-6 hrs.in a shaking incubator and incubated at 30℃ for 2 days. The confirmation and appearance of most diploids were again confirmed with a light microscope, and the 100µl of culture was inoculated overnight to make competent cells. The competent cells were transformed with either an empty vector or a triple-sgRNA CRISPR reagent pCas9-sgRNA^{511B,USERX,FGF20}. For the most part, we followed a *GFP* monster protocol. However, we did not use haploid or diploid-specific selection (Suzuki et al., 2012). Instead, CRISPR selection was sufficient to enrich unique combined genotypes.

2.4.23. Genomic DNA isolation, whole genome sequencing, and SNP analysis

Genomic DNA was purified from single-humanized Hs*a*1 and quintuple-humanized *Hsα1,α2,α3,α4,α7* strains using the Monarch Nucleic acid purification kit according to the manufacturer's protocol (NEB). Spheroplasts were obtained before the genomic DNA extraction for a high-quality DNA prep. Wild type and engineered strains were sequenced using Illumina MiSeq 2x150 at 30x coverage using 150-bp paired-end reads. The Geneious Pro Software (Kearse et al., 2012) and its included tools were utilized for pairing paired-end sequences, trimming ends and adapters based on quality using the BBDuk tool, and reference mapping using the Geneious Read Mapper algorithm at medium-low sensitivity, iterated 5 times. Reads were mapped to a reference BY4741 strain (for the engineered strain, the same reference sequence was used, but replacing the sequences of the 5 engineered genes with their human ortholog to ensure alignment

with the humanized loci reads) using the Geneious Prime Software 2020.2.4 algorithm. Mapping was run at medium sensitivity: word length to allow for matching was set to 18 with a maximum permitted mismatch of 20% of the read length, maximum mismatches per read were set to 20%, a minimum 80% overlap was required, and reads with errors were set to accurately be mapped to repeat regions - this was iterated 5 times to give the final mapping. For analysis, low coverage regions of below 2 standard deviations from the mean were excluded from SNP-calling, and only SNPs with a variant frequency of 0.90 or higher were considered. SNPs were called with a minimum variant frequency of 0.25, and low coverage regions below 2 standard 1109 deviations from the mean were excluded. SNPs that were unique to the engineered strain (i.e. not in the mappings of the wild-type strain) were marked. To ensure SNPs were not introduced by off-target CRISPR effects, 300bp up- and down-stream of each SNP was searched for CRISPR-gRNA alignment at 75% sequence similarity.

2.4.24. Construction of yeast expression vectors harboring human constitutive- and immuno-βs

Human β subunits, with flanking Type IIS restriction sites generating unique overhangs, were obtained as gBlocks from IDT. To obtain a yeast expression vectors for native expression of 7 human constitutive- (pCN7) and immuno-βs (pIN7), we PCR amplified 5'UTRs (~500bp) and 3'UTRs (`250bp) of the orthologous yeast genes using genomic DNA of BY4741 strain as a template. Each 5'UTR, gBlock (human β ortholog) and 3'UTR were cloned as transcription units (TUs) using Type IIS (Golden Gate)-based cloning. The TUs were sequence-verified followed by cloning 7 constitutive- and immuno-β TUs in yeast vector (*CEN6, URA*+). To obtain a yeast expression vectors for heterologous expression of 5 functionally non replaceable human

constitutive- (pCH5) and immuno-βs (pIH5) (Kachroo et al. , 2015; Laurent et al. , 2020), the human β gBlocks were cloned as transcription units (TUs) using strong yeast promoters and terminators from MoClo toolkit and Type IIS (Golden Gate)-based cloning (Lee et al. , 2015). The TUs were sequence-verified followed by cloning 5 constitutive- and immuno-β Tus in yeast vector (*CEN6, URA+*).

2.4.25. Proteomic Sample Preparation

Singly humanized Hs*a*1, quintuple-humanized *Hsα1,α2,α3,α4,α7* and sextuple humanized *Hsα1,α2,α3,α4,α7,β3* yeast strains were transformed with empty vector or plasmids expressing 7 human constitutive- (pCN7) or immuno- βs (pIN7). The cell pellet obtained from the strains was resuspended in Digestion Buffer (50mM Tris, 2mM $CaCl_2$) and lysed by bead beating with glass beads for 3 x 1 min. The resulting whole-cell lysate was recovered and mixed 1:1 with 2,2,2-trifluoroethanol. Samples were reduced by incubation with 5 mM tris - carboxyethyl)phosphine (TCEP solution, Pierce) at 60°C for 40 min. Reduced samples were then alkylated by incubation with 15 mM iodoacetamide at room temp. for 30 min. 7.5mM Dithiothreitol (DTT) was added to quench excess iodoacetamide. Samples were then diluted 10-fold with Digestion Buffer and 2 µg trypsin was added for digestion. Samples were incubated at 37°C for 5 hr. Tryptic digestion was quenched with 1% formic acid, and samples were concentrated by vacuum centrifugation to reduce volume to less than 300 µl. Digests were cleaned using HyperSep C18 SpinTips (Thermo) according to manufacturer's protocol. Eluted peptides were briefly dried by vacuum centrifugation, then resuspended in 5% acetonitrile, 0.1% formic acid.

2.4.26. LC-MS/MS Analysis

Tryptic peptides were separated by reverse phase chromatography on a Dionex Ultimate 3000 RSLCnano UHPLC system (Thermo Scientific) with an Acclaim C18 PepMap RSLC column using a 3-42% acetonitrile gradient over 60 min. Peptides eluted directly into a Thermo Orbitrap Lumos mass spectrometer by nano-electrospray. Data-dependent acquisition was applied, with precursor ion scans (MS1) collected by FTMS at 120,000 resolution and HCD fragmentation scans (MS2) collected in parallel by ITMS with 3-second cycle times. Monoisotopic precursor selection and charge-state screening were enabled, with ions > +1 charge selected. Dynamic exclusion was applied to selected ions +/- 10 ppm for 30 s. Raw data was processed using Proteome Discoverer 2.2 (Thermo Scientific). Mass spectra were searched against a protein sequence database comprising the *Saccharomyces cerevisiae* reference proteome (UniProt OX: 559292), human proteasome proteins, and a list of common protein contaminants (MaxQuant). Searches were restricted to fully tryptic peptides only, allowing up to two missed cleavages. A precursor tolerance of 5 ppm and fragment mass tolerance of 0.5 Da were used. Static modifications of carbamidomethyl cysteine and dynamic modifications of oxidized methionine and protein N-terminal acetylation and/or methionine-loss were considered. High-confidence peptide-spectrum matches (PSMs) were filtered at a false discovery rate of <1% as calculated by Percolator. Peptide abundances based on extracted ion-chromatography (XIC) feature intensities were calculated using the Label-Free Quantitation (LFQ) workflow. Due to conservation of sequence between human and yeast proteasome proteins, peptide and protein abundances were calculated using an in-house workflow to remove degenerate peptides. Isoleucine/leucine sequence variants were collapsed into single peptide groups. For each peptide, protein groups generated by Proteome Discoverer 2.2 were checked to make sure all peptides used were unique to either yeast or human protein. Protein abundances were calculated as the sum of all unique peptides matching to those proteins. Only

proteins detected in two separate injections of at least one sample were included in the final dataset. The consolidated data is provided as Table S2.2.

2.4.27. Microscopy

The yeast cells were grown to mid-log phase in selective media to maintain plasmids and were imaged with DMi6000B microscope (Leica Microsystems). Cell size measurements were performed with a minimum 50-100 cells per strain using FIJI/ImageJ.

2.5. Supplementary Information

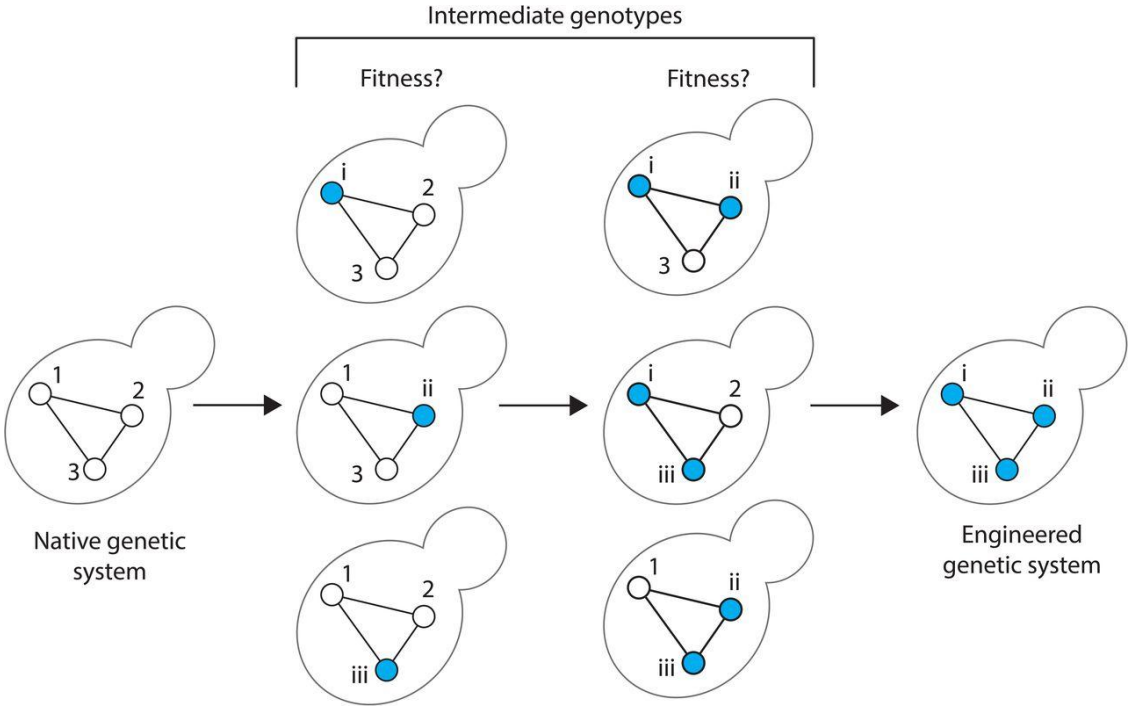


Figure S2.1. Multiple routes to the engineering of biological systems using a sequential replacement strategy.

Let's examine a simple 3-gene pathway. A sequential replacement strategy to fully engineer this module will navigate 6 possible intermediate genotypes. The following formula can calculate the number of genotypes: $nCr = (n!/n - r!)\ (r!)$, where n represents a total number of genes to be combined (for example, depicted above, n = 3). And r represents the combined intermediates, like for a single-gene intermediates r = 1 and similarly for a 2-gene intermediates r = 2. Furthermore, while the strategy explores 6 intermediate genotypes, the distinct routes to complete pathway engineering are 9 as some genotypes may require gene 1 and then gene 2, or gene 2 and then gene 1. The

higher the number of genes to be engineered, the more the number of intermediate genotypes to be explored. As a reference, for the humanization of the heptameric α proteasome core, a total of 126 intermediate hybrid yeast-human genotypes exist towards humanizing the entire yeast α core.

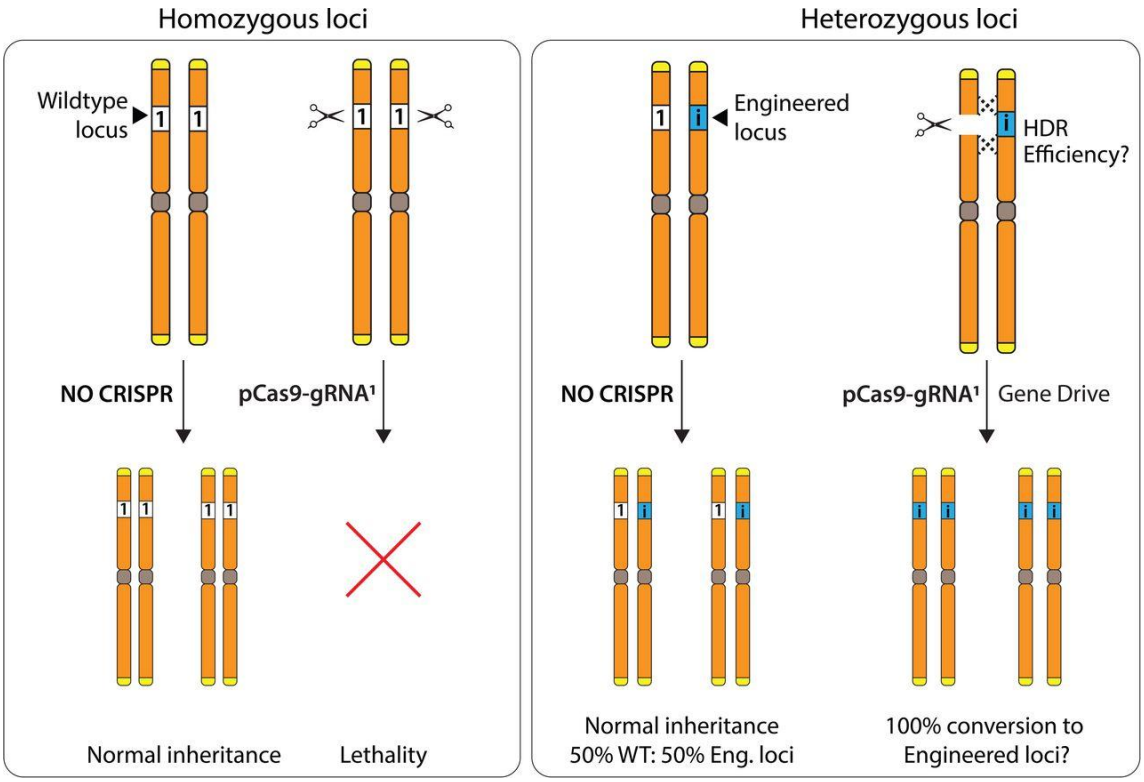


Figure S2.2. MERGE enables the highly efficient conversion of heterozygous to homozygous loci.

Every locus in yeast follows a normal (1:1) inheritance pattern of segregation. However, CRISPR reagent targeted to a haploid or a diploid locus leads to lethality. Similarly, a heterozygous yeast locus follows 1 (1, wild-type) : 1 (i, engineered) inheritance pattern. CRISPR reagent targeted to one of the heterozygous alleles initiates homologous recombination using the CRISPR-resistance homologous chromosome as a repair template, thus enabling efficient conversion to homozygous loci, mimicking a gene drive. The subsequent segregation pattern of 4 : 0 allows propagation of only desired loci.

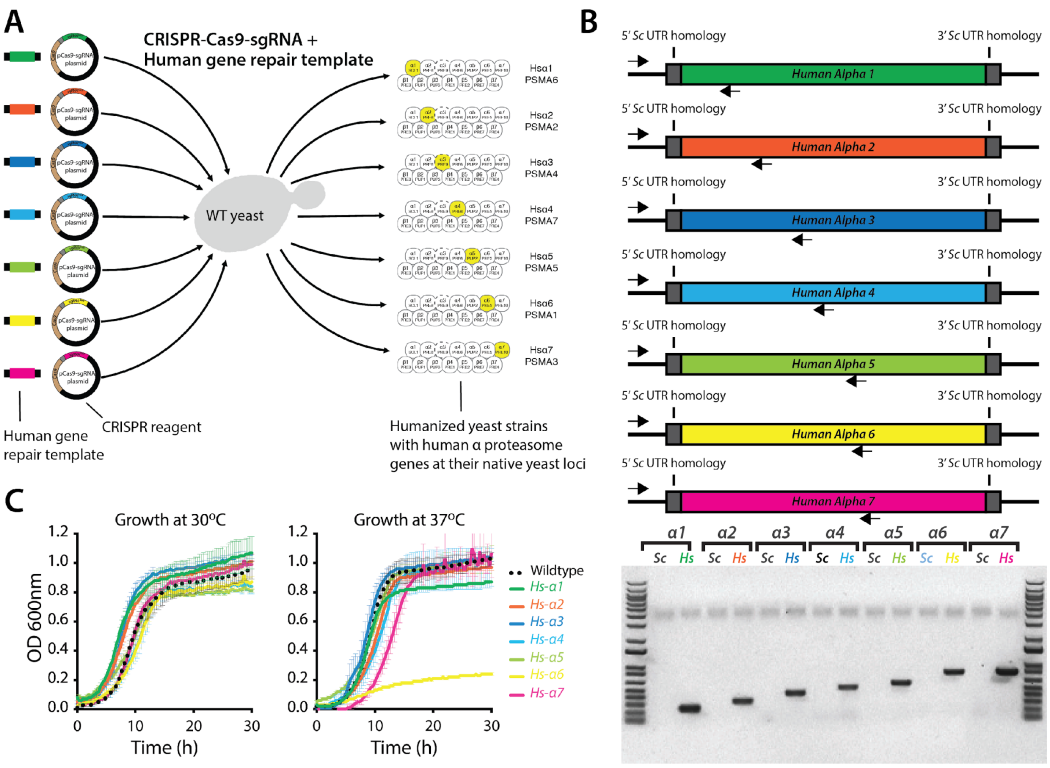


Figure S2.3. General scheme of CRISPR-Cas9-based gene replacement.

(A) For each gene, a 'knockout' (KO) plasmid (CRISPR reagent) is constructed that expresses the Cas9 endonuclease and a sgRNA with a targeting sequence to the gene of interest. To replace the gene of interest, the CRISPR reagent is co-transformed into yeast with a linear repair DNA, consisting of the replacing gene sequence flanked by long (60-200bp) homology sequences to the flanking region of the genome. Cas9 and the sgRNA are co-expressed and target the gene of interest for cutting, allowing repair by homologous recombination (HR) with the provided repair DNA. (B) After obtaining transformants, sequencing across the junction reveals successful replacement (as shown for the human α proteasome subunits replaced at the native yeast loci). PCR for the yeast locus (*Sc*) is negative, whereas for a human gene (*Hs*) is positive. (C) The growth of singly humanized α-subunits is not significantly impaired. Each curve represents an averaging of 4 replicates for each indicated strain, grown in YPD for 24 hrs. No strains exhibit significantly impaired growth, though strains with human subunits α4, 5, and 6 show some extension of lag phase at 30°C and marginally slowed doubling times at 30°C. The behavior is similar at 37°C except in *Hs-α6,* which exhibited a significant growth defect.

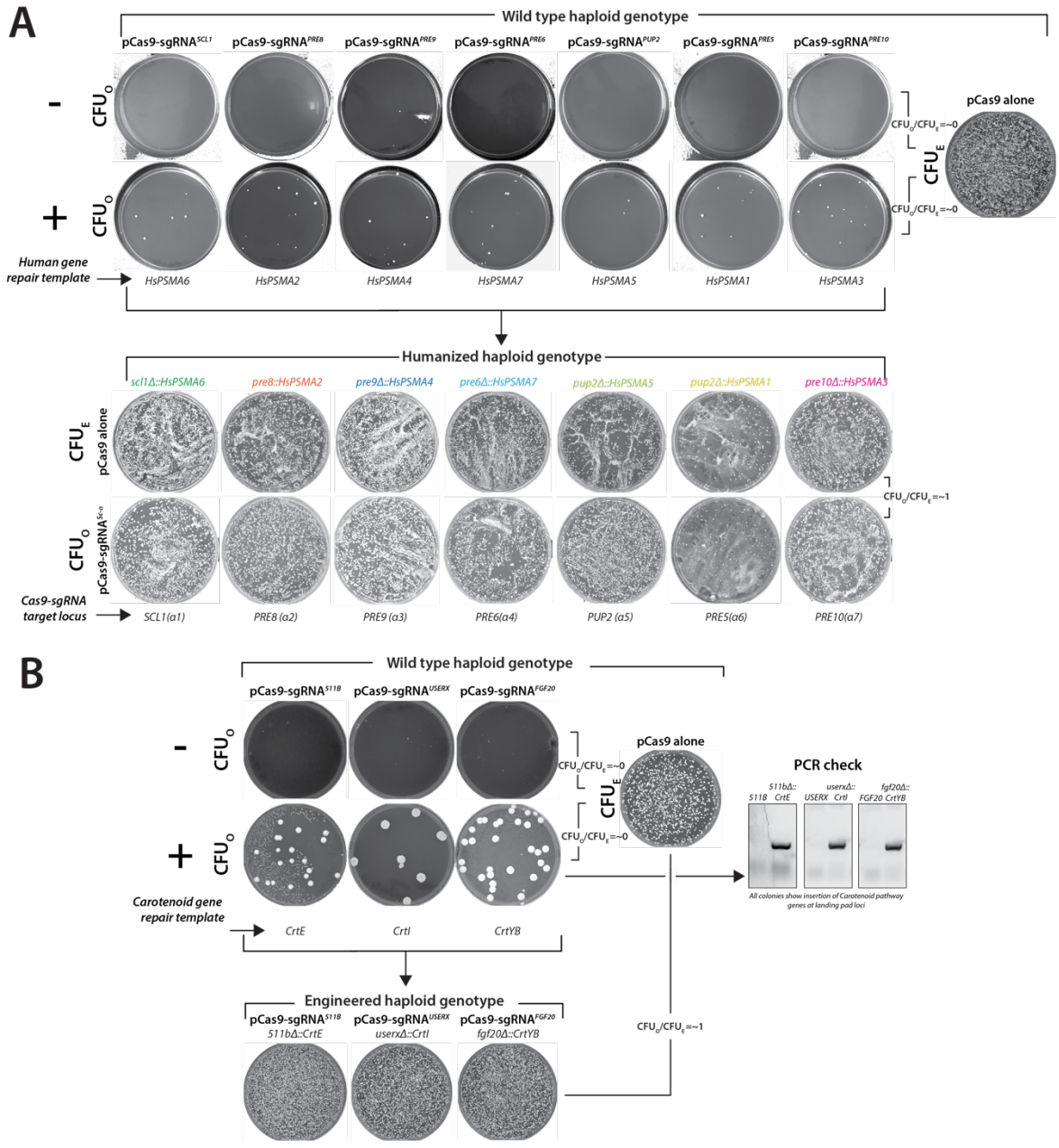


Figure S2.4. General scheme of CRISPR-Cas9-based gene replacement, CRISPR-mediated sensitivity, and resistance of yeast loci.

(A) CRISPR reagent (pCas9-sgRNA^{locus}) targeted to any yeast locus often leads to lethality with few surviving colonies (CFU_O). The corresponding transformation of a control plasmid (pCas9 alone) estimates the transformation efficiency (CFU_E). CRISPR reagents targeting 7 yeast α proteasome genes show lethality ($CFU_O/CFU_E = \sim 0$) with or without the repair template. After humanizing the corresponding yeast loci, each single-humanized haploid strain shows resistance to further targeting by the corresponding CRISPR reagent, respectively ($CFU_O/CFU_E = \sim 1$). (B) CRISPR reagents

targeted to landing pad loci (*511B*, *USERX* and *FGF20*) enabled editing by introducing carotenoid gene transcription units *CrtE*, *CrtI* and *CrtYB* (provided as PCR fragment repair templates). The engineered strains show resistance to further cutting by the corresponding CRISPR reagents ($CFU_O/CFU_E = \sim 1$). PCR-based genotyping confirmed the integration of each carotenoid transcription unit for every colony tested.

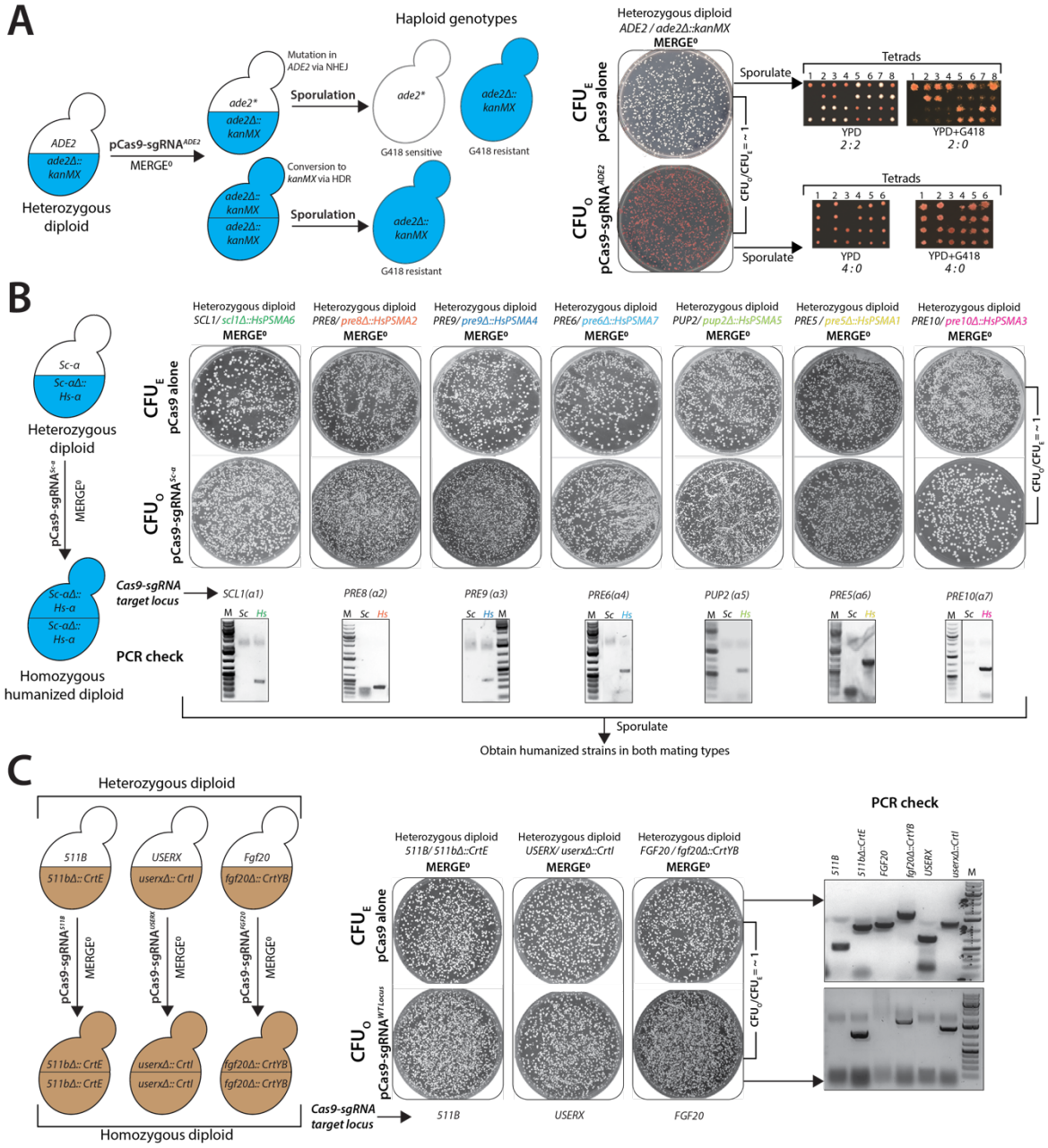


Figure S2.5. MERGE⁰ enables the efficient conversion of single heterozygous to homozygous loci.

(A) A schematic of possible genotypes after the *ADE2* locus in a heterozygous diploid *ADE2*/*ade2Δ::kanMX* strain is subjected to a CRISPR-Cas9-mediated DSB followed by sporulation. The repair via NHEJ could mutate the locus (*ade2**) or repair via HDR converts the locus to *ade2Δ::kanMX*, either scenario leads to a red colony phenotype. However, post sporulation, the prior scenario will show only 50% G418 resistant colonies compared to the latter outcome, which will exhibit 100% G418 resistance. Tetrad dissection of pooled cells from heterozygous diploid strains before MERGE^0 (pCas9 alone) or after MERGE^0 (pCas9-sgRNA^{ADE2}) shows that before MERGE^0 cells appear as 2 : 2 (red : white) and 2 : 2 (G418- resistant : G418-sensitive) phenotype. However, after MERGE^0, cells show 4 : 0 (red : white) phenotype. Furthermore, all 4 spores in the post MERGE^0 strain show 4 : 0 G418 resistance. (B) Schematic shows CRISPR reagent targeted to one of the heterozygous alleles (wild type yeast gene) in a diploid strain permits efficient conversion to homozygous humanized loci. MERGE^{0}efficiently converts wild-type yeast loci α proteasome genes to humanized loci ($CFU_O/CFU_E = \sim 0$). PCR-based genotyping verified the conversion to the CRISPR-resistant humanized loci for every colony tested. (C) Schematic shows how CRISPR reagent targeted to the landing pad loci in a heterozygous diploid enables conversion to engineered carotenoid loci. MERGE^0 converts heterozygous landing pad loci (*511B, USERX,* and *FGF20*) to engineered carotenoid loci ($CFU_O/CFU_E = \sim 1$). PCR-based genotyping shows heterozygous status of the loci before MERGE^0 and conversion to engineered carotenoid loci after MERGE^0. Sporulation yields haploids of both mating types with only engineered loci.

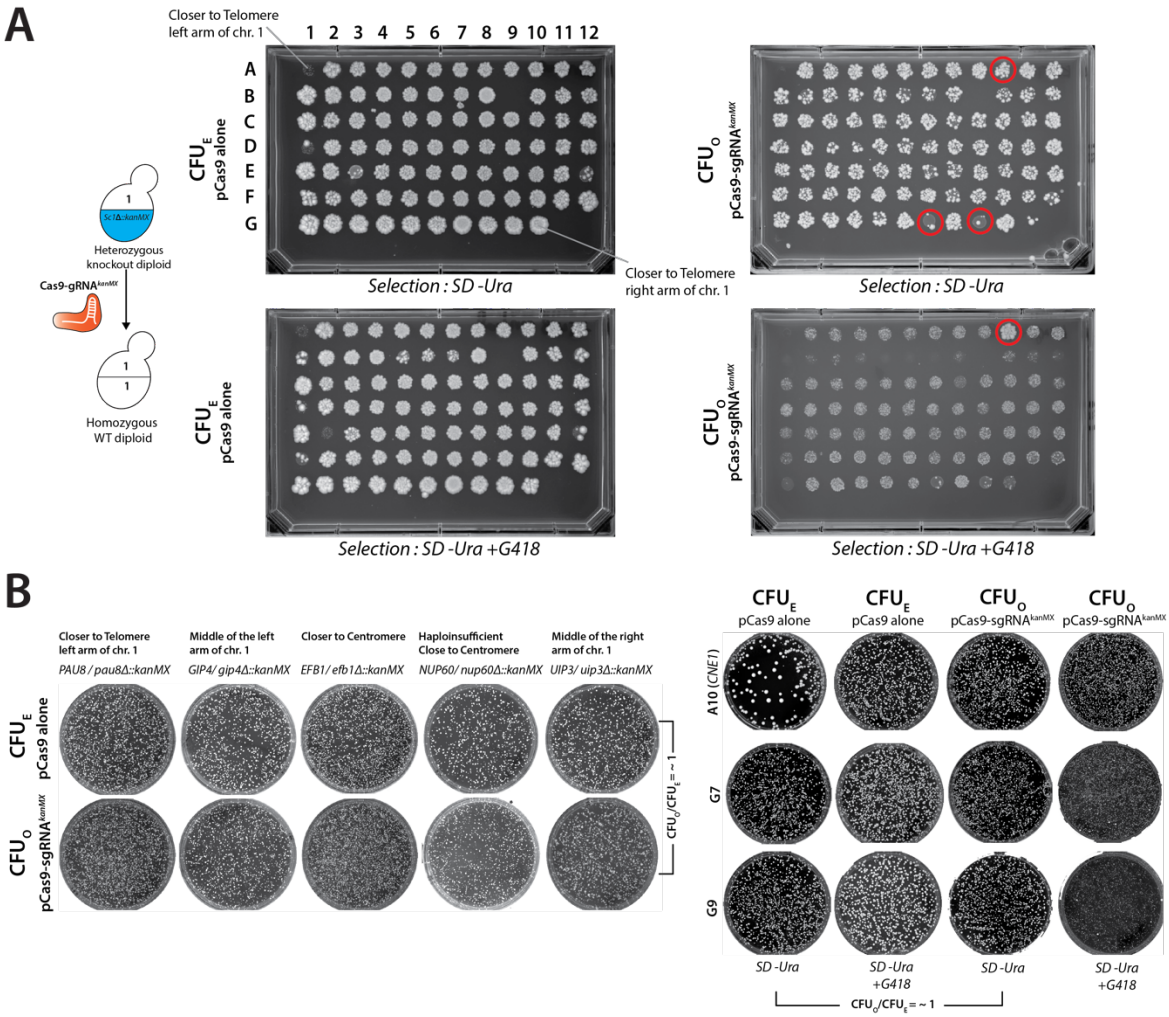


Figure S2.6. MERGE^0 is efficient at converting many yeast heterozygous to homozygous loci independent of the position on Chromosome 1.

(A) Schematic shows a CRISPR reagent (pCas9-sgRNA^{kanMX}) targeted to the *kanMX* allele in a hetKO strain converts the locus to homozygous wild-type. The hetKO strains were arrayed as genes close to the left-arm of a telomere through the centromere to the right arm of a telomere. The strains were transformed with pCas9 alone (CFU_E) or with pCas9-sgRNA^{kanMX} (CFU_O) and spotted on either SD-Ura or SD-Ura with G418 selection. Every genotype shows spots with a similar density of colonies showing no lethality due to the CRISPR reagent ($CFU_O/CFU_E = \sim 1$). The spots on SD-Ura with G418 selection show G418 resistance in the case of pCas9 alone transformants. However, all but one (A10 - *CNE1*) strain transformed with pCas9-sgRNA^{kanMX} lost the *kanMX* cassette resulting in G418 sensitivity. (B) MERGE^0 mediated conversion of loci was tested on several unique hetKO genotypes on Petri plates, demonstrating agreement with 96-well spot assays. All tested strains show no lethality due to the CRISPR reagent ($CFU_O/CFU_E = \sim 1$). Ambiguous spots on a 96-well plate (A10, G7 and G9) were further analyzed on Petri plates. G7 and G9 spots show no lethality due to CRISPR reagent ($CFU_O/CFU_E = \sim 1$) and loss of *kanMX* cassette. While the A10 (*CNE1*) strain is resistant to the CRISPR reagent, the cells retain G418 resistance. It is unclear why this behavior occurs.

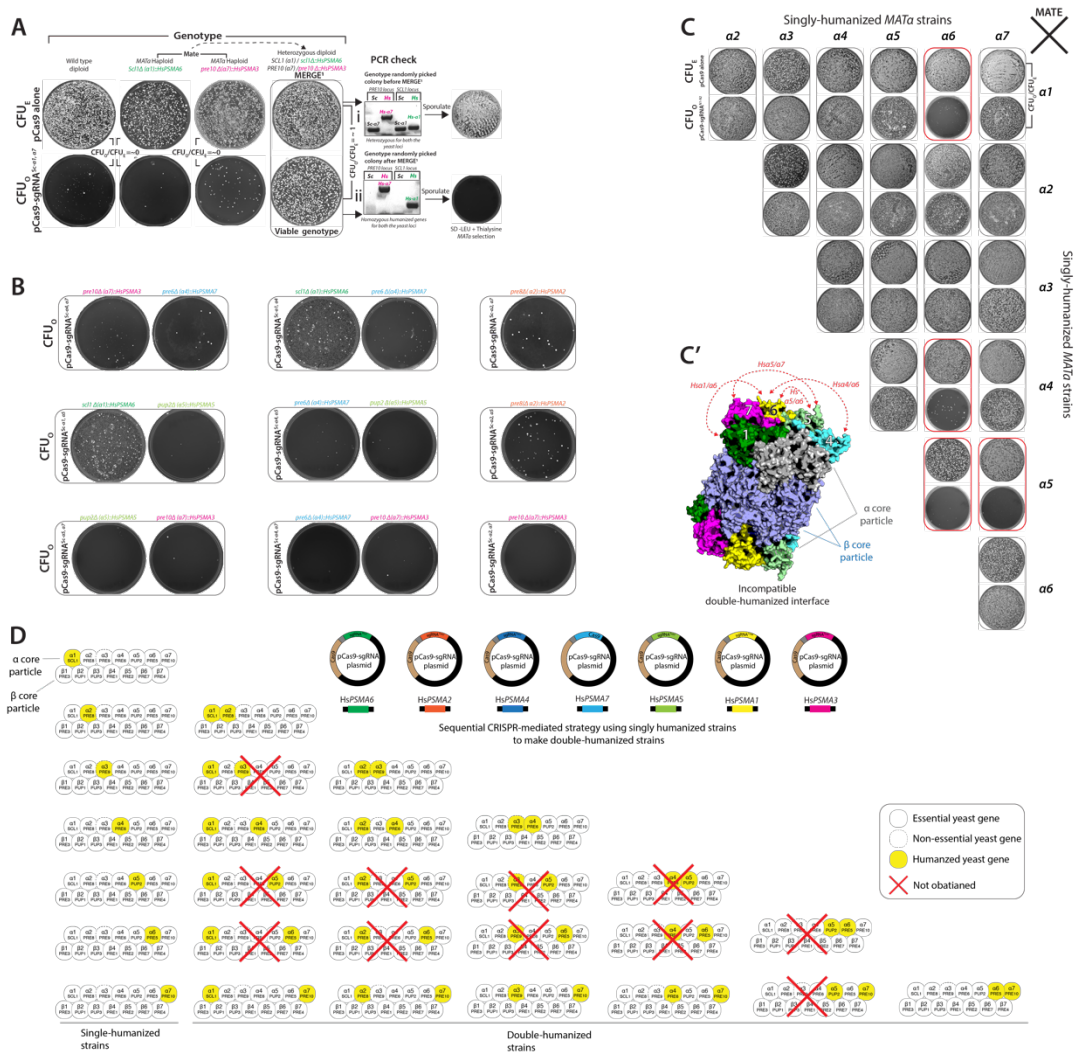


Figure S2.7. MERGE^0 is efficient at simultaneously combining two heterozygous to homozygous loci.

(A) Double-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\alpha1,\alpha7}$) targeted to two wild type yeast loci (proteasome α1 and α7 genes) quantifies the efficiency of MERGE^1. The transformation of pCas9-sgRNA$^{Sc\alpha1,\alpha7}$ in the wild type, and single-humanized Hsα1 or Hsα7 strains is lethal, suggesting ON-target activity (CFU_O/CFU_E = ~0). However, the transformation of pCas9-sgRNA$^{Sc\alpha1,\alpha7}$ in diploid heterozygous humanized strain shows no lethality (CFU_O/CFU_E = ~1). PCR-based genotyping of surviving colonies after MERGE^1 shows conversion of both yeast to the humanized loci compared to before MERGE^1. (B) All tested single-humanized strains transformed with double-sgRNA CRISPR reagents show lethality, demonstrating that selection only allows survival of viable paired genotypes. (C)Mating each single-humanized strain of opposite mating-types (obtained via MERGE^0) provided 21 heterozygous genotypes to test MERGE^1. Heterozygous diploid humanized strains transformed with double-sgRNA CRISPR reagents specific to each yeast allele allow testing viability of all paired-humanized genotypes (CFU_O). If the genotype is permitted, the strategy enables survival similar to the pCas9 alone transformation (CFU_E). While the majority of combined double-humanized genotypes are viable (CFU_O/CFU_E = ~1), specific genotypes (*Hsα1α6*, *Hsα4α6*, *Hsα5α6* and *Hsα5α7*) were inviable, suggesting incompatible genotypes. Other combinations, such as those involving the genotypes with Hsα2, showed a slow-growth phenotype. (C') Yeast proteasome structure (PDB-1RYP) with α and β cores show the

humanized *α* subunits that are incompatible as paired genotypes. Except for *α5* and *α6*, the subunits are missing neighboring interactions within the α core. (D) A sequential strategy using single-humanized strains acquired several pairwise combinations of human α subunits. However, several other genotypes could not be obtained, revealing the drawback of the method. MERGE[1] is far more efficient in identifying incompatible paired genotypes

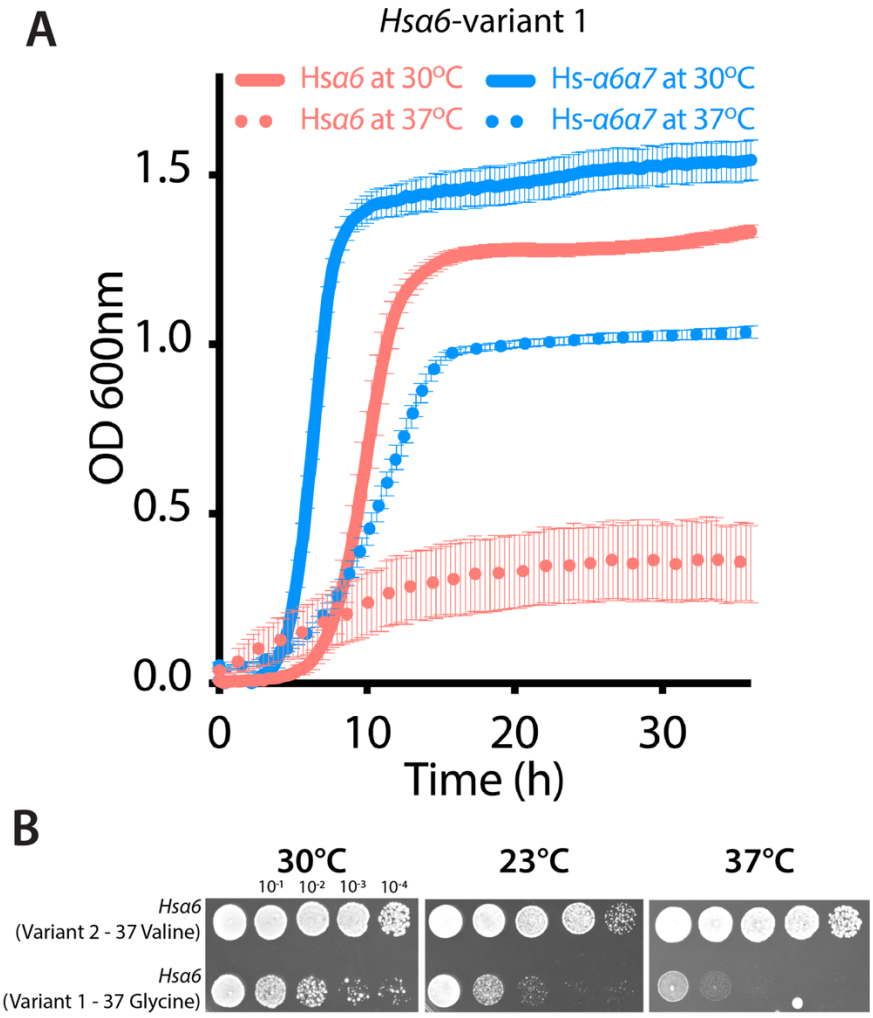


Figure S2.8. The variant-specific phenotype of single-humanized Hs*α6* (*PSMA1*) yeast strain.

(A) A growth assay performed on a single-humanized *Hsα6* strain (Variant 1, amino acid residue 37>Glycine) shows a temperature-sensitive (TS) phenotype at 37°C (dotted red line) compared to growth at 30°C (solid red line). The combined genotype of Hsα6,*α*7 partially rescues the TS phenotype at 37°C (dotted blue line) compared to the growth at 30°C (solid blue line). (B) Spotted dilutions of single-humanized *Hsα6* strains (Variant 1, amino acid residue 37>Glycine) and (Variant 2, amino acid residue 37>Valine) show Variant 2 with no growth defect at 30°C, 23°C and 37°C compared to Variant 1.

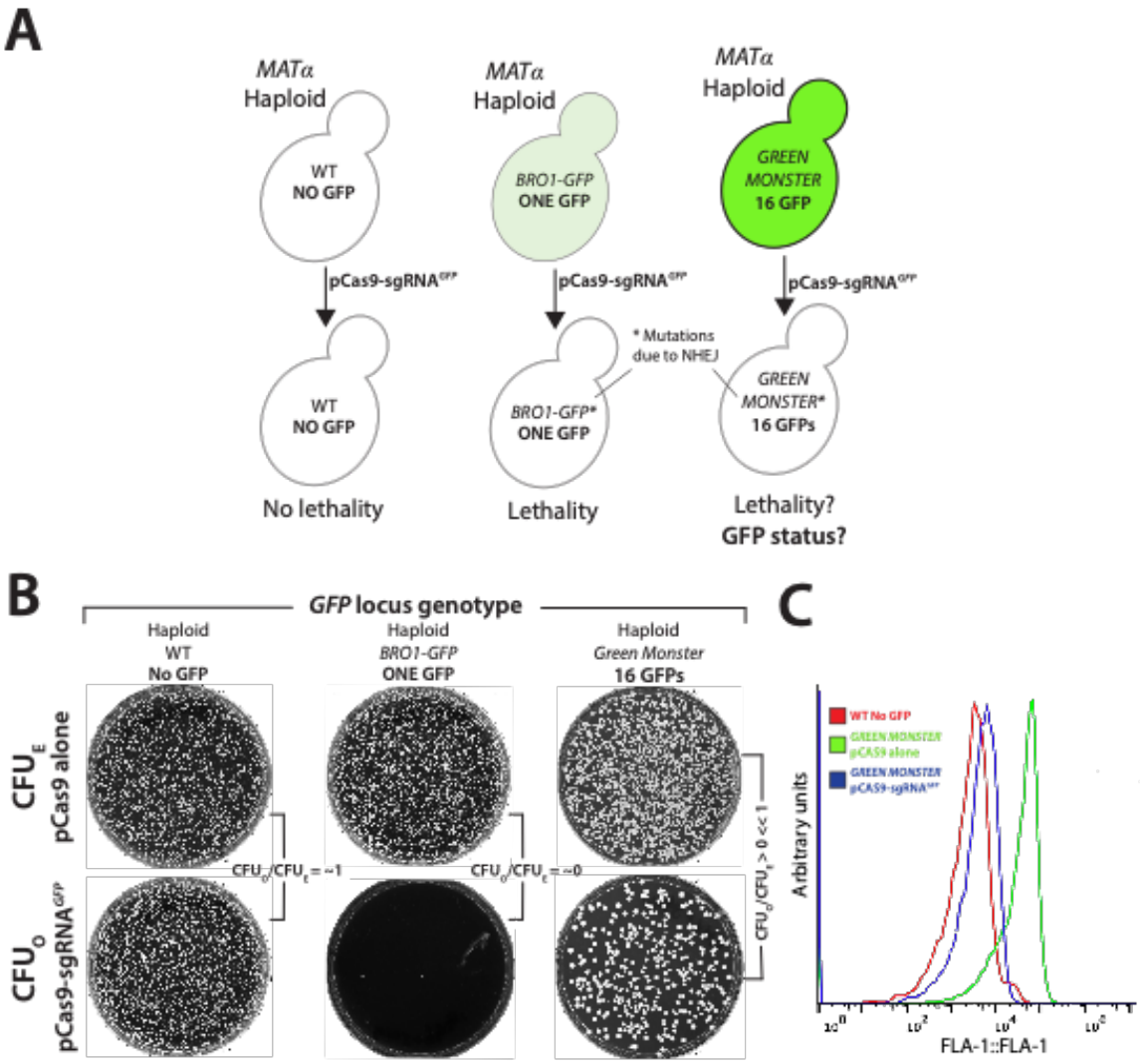


Figure S2.9. MERGE is scalable to target multiple yeast loci simultaneously.

(A) The schematic shows a CRISPR-based strategy targeting one or many genetic loci in yeast. Transformation of a CRISPR reagent (pCas9-sgRNA^{GFP}) targeting a *GFP* cassette in a Green Monster strain tests the scalability. Transformation in the wild type (no *GFP*) strain should show no lethality, whereas strains harboring single or multiple *GFP*s should show a lethal phenotype. (B) The pCas9-sgRNA^{GFP} transformation in wild type yeast showed viable cells ($CFU_O/CFU_E = \sim 1$). In contrast, transformation in a strain harboring a single *GFP* cassette was lethal. While the Green Monster strain (*GFP* monster, 16 *GFP* cassettes) showed more viable cells than the single-*GFP* strain upon transformation with pCas9-sgRNA^{GFP}. (C) Using flow cytometry, the pooled mixture of yeast cells that survive pCas9-sgRNA^{GFP} showed significantly reduced *GFP* expression compared to the pCas9 alone transformed cells, suggesting that nearly every *GFP* cassette was successfully targeted and mutated via NHEJ.

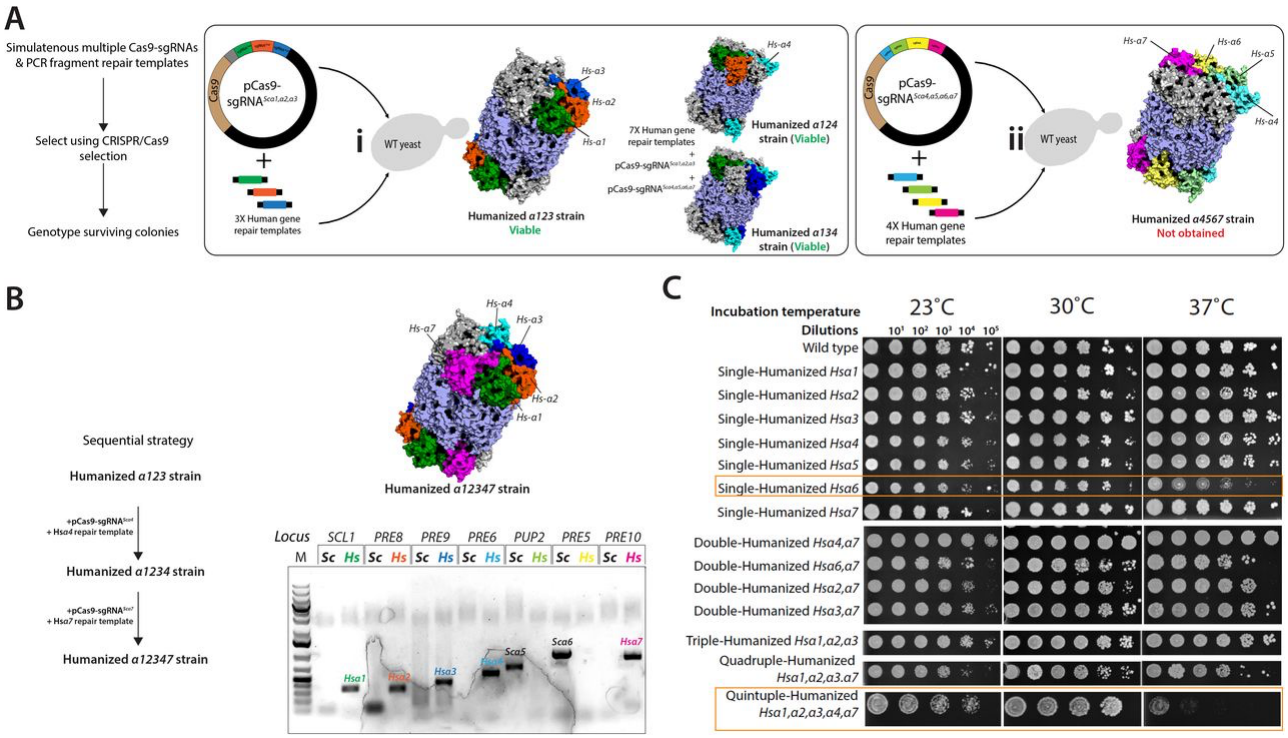


Figure S2.10. Scheme for obtaining triple-, quadruple-, quintuple- and sextuple-humanized α-subunit strains *via* a sequential strategy.

(A)Two multiple-sgRNA CRISPR plasmids were constructed, one expressing sgRNAs targeting yeast α subunits 1, 2, and 3 (pCas9-sgRNA$^{Sca1,\alpha2,\alpha3}$), and the other expressing sgRNAs targeting α subunits 4, 5, 6, and 7 (pCas9-sgRNA$^{Sca4,\alpha5,\alpha6,\alpha7}$). They were transformed individually with the respective human gene repair DNA. A single colony with triple-humanized *Hs-α1α2,α3* (i) was obtained. However, no clone for a quadruple humanized *Hs-α4α5,α6,α7* (ii) could be obtained. A similar strategy also generated triple-humanized *Hs-α1α2,α4* and *Hs-α1α3,α4* genotypes. (B) Sequential strategy generated a quintuple-humanized *Hs-α1α2,α3,α4,α7* strain confirmed via locus-specific PCR showing the presence of human and the absence of yeast loci. Sanger sequencing and whole-genome sequencing confirmed the genotype. (C) Spotted dilutions of double-, triple-, quadruple-, quintuple- and sextuple humanized *α* proteasome strains at 23°C, 30°C and 37°C show singe-humanized Hs-*α6* (variant 1) and quintuple-humanized *Hs-α1,α2,α3,α4,α7* strain with a temperature-sensitive growth phenotype.

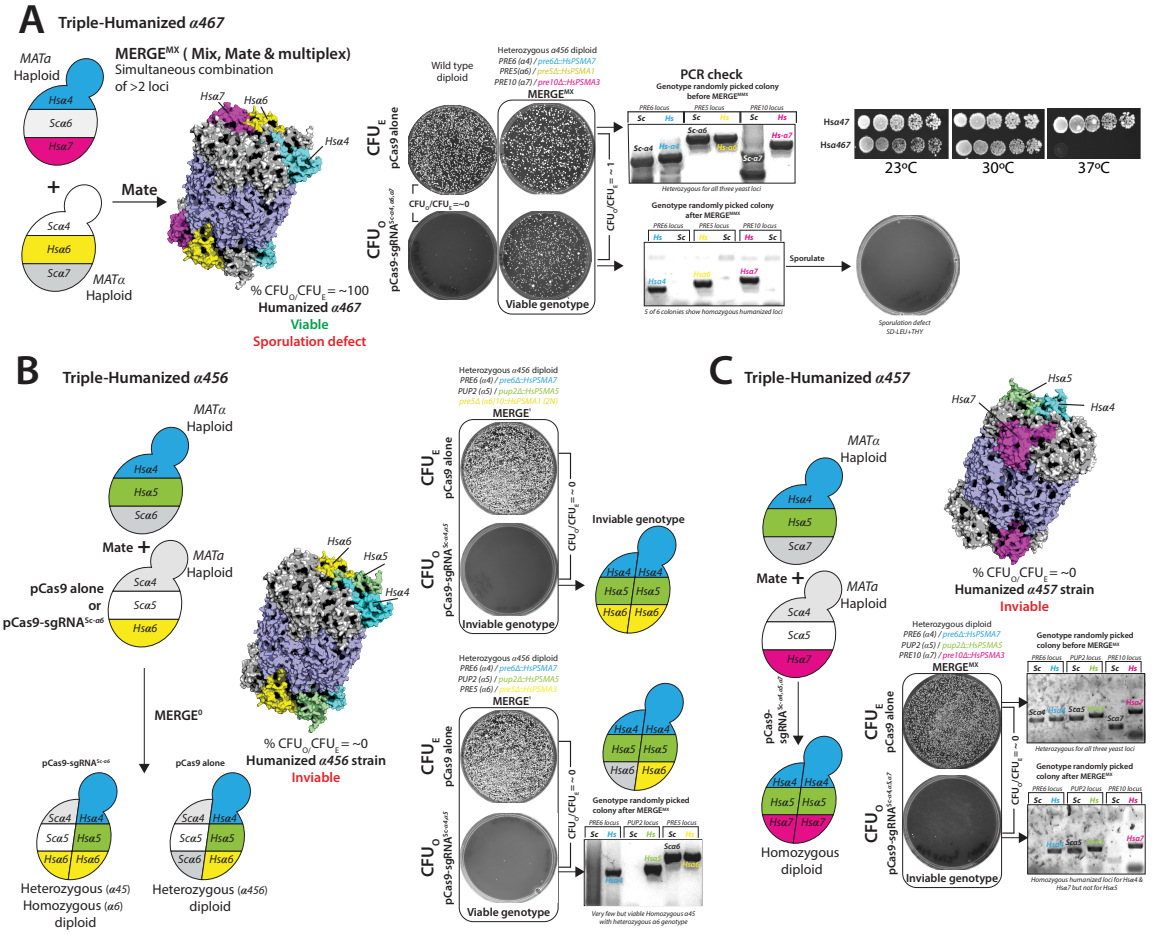


Figure S2.11. MERGE tests the combination of several triple-humanized yeast α-subunit genotypes.

(A) MERGE^{MX} (Mate and multiplex) strategy combined 3 loci after mating a strain with 2 humanized loci (Hs*α*4,*α*7, *MAT a*) with a strain carrying 1 humanized locus (Hs*α*6, *MATα*) (biological replicate of data shown in Figure 4E). The transformation of pCas9-sgRNA$^{Sc\alpha4,\alpha6,\alpha7}$ in a mated mix selected a diploid triple-humanized strain (CFU_O/CFU_E = ~1). PCR-based genotyping of several colonies after MERGE^{MX} shows conversion of all yeast to the humanized loci (5 of 6). Spotted dilutions of humanized *Hs-α4,α7* and *Hs-α4,α6,α7* strains at 23°C, 30°C and 37°C show the triple-humanized strain with a temperature-sensitive phenotype at 37°C. The strain does not yield viable haploid spores observed as no growth on haploid-specific selection media (SD-LEU+Thialysine). (B) Double-humanized *Hs-α4α5* (haploid *MATα*) and single-humanized *Hs-α6* (haploid *MATa*, harboring a pCas9 alone or pCas9-sgRNA$^{Sc-\alpha6}$ vector) strains were mated, generating two distinct heterozygous diploid strains, one heterozygous for three loci and another for two with homozygous *Hs-α6* locus. Each strain was transformed with double-sgRNA CRISPR plasmid (pCas9-sgRNA$^{Sc\alpha4,\alpha5}$) to perform MERGE[1]. The transformation caused lethality in both strains (CFU_O/CFU_E =~0). However, a few survivors in the context of heterozygous *Hs-α6* locus showed the conversion of two loci (*Hs-α4* and *Hs-α7*), suggesting an incompatible triple-humanized genotype but viable double-humanized genotype harboring a heterozygous yeast-human *α6* locus. **(C)** Double-humanized haploid *Hs-α4α5* (*MATα*) and *Hs-α7* (*MATa* a) strains were mated. The resulting heterozygous diploid strain was transformed with a triple-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\alpha4,\alpha5,\alpha7}$). The transformation caused lethality with few survivors (CFU_O/CFU_E =~0). Genotyping of the survivors showed the conversion of two 2 of 3 loci (*Hs-α4* and *Hs-α7*), suggesting an incompatible triple-humanized genotype.

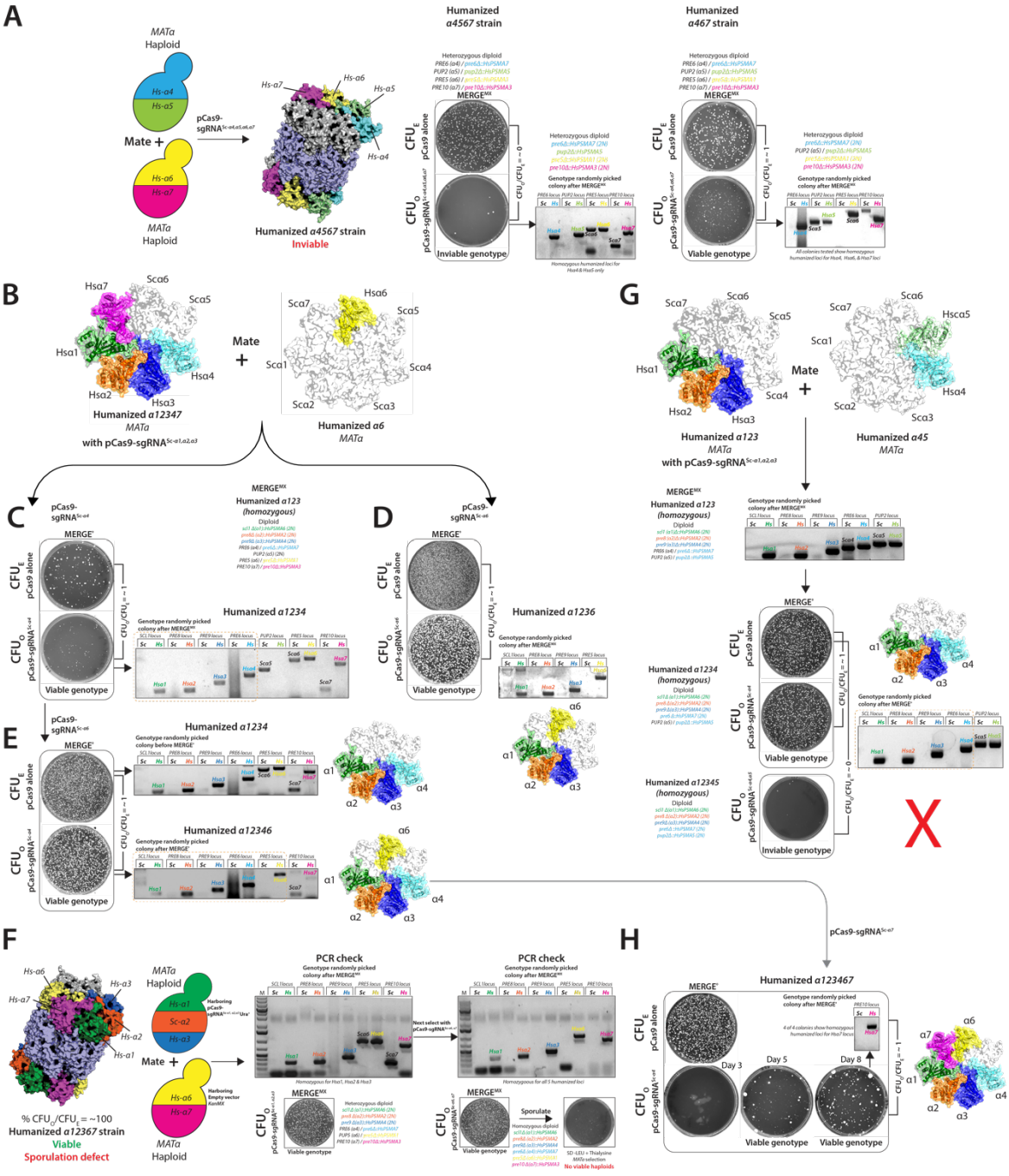


Figure S2.12. MERGE efficiently tests several higher-order combinations (>3) of humanized yeast α-proteasome core.

(A) Double-humanized haploid *Hs-α4α5* (*MATα*) and *Hs-α6α7* (*MATα*) strains were mated. The resulting heterozygous diploid strain was transformed with quadruple-sgRNA CRISPR reagent (pCas9-sgRNA^{Scα4,α5,α6,α7}) to perform MERGE^{MX}. The transformation caused lethality with few survivors (CFU_O/CFU_E =∼0). The survivors showed the conversion of only 2 of 4 loci (*Hs-α4* and *Hs-α5*), suggesting an incompatible quadruple-humanized genotype. However, the transformation of triple-sgRNA CRISPR reagent (pCas9-sgRNA^{Scα4,α6,α7}) in this strain was viable (CFU_O/CFU_E =∼1). The genotyping of randomly picked colonies showed the conversion of 3 yeast to human loci (*Hs-α4, Hs-α6,* and *Hs-α7*), suggesting a triple-humanized genotype is viable with a heterozygous yeast-

human *α5* locus. (B) Mating quintuple-humanized haploid *Hs-α1α2α3α4α7* [*MATa*, also harboring a triple-sgRNA CRISPR reagent (pCas9-sgRNA^{Scα1,α2,α3})] and single-humanized *Hs-α6* (*MAT α*) strains simultaneously combined three yeast to humanized *Hsα1, Hsα2, Hsα3* loci in a diploid yeast (CFU_O/CFU_E =∼1). The strain harbors heterozygous yeast-human alleles for *α4, α6,* and *α7* loci. Transformation of the diploid strain with (C) pCas9-sgRNA^{Scα4} or (D) pCas9-sgRNA^{Scα6} generated viable quadruple-humanized strains (Hs-*α1α2α3α4* and Hs- *α1α2α3α6*) (CFU_O/CFU_E =∼1). Genotyping of randomly picked colonies confirmed the homozygosity of the loci. (E) Using the quadruple-humanized strain *Hs-α1α2α3α4* as a background, the transformation of pCas9-sgRNA^{Scα6} yielded a viable quintuple-humanized (Hs- *α1α2α3α4α6*) genotype (CFU_O/CFU_E =∼1) confirmed using locus-specific PCR. (F) Double-humanized haploid *Hs-α6α7* (*MATα*) and triple-humanized *Hs-α1,α2,α3* [*MATa*, also harboring a triple-sgRNA CRISPR reagent (pCas9-sgRNA^{Scα1,α2,α3})] strains were mated. MERGE^{MX} simultaneously combined three yeast to humanized *Hsα1, Hsα2, and Hsα3* loci in the diploid background (CFU_O/CFU_E=∼1). The next transformation, using pCas9-sgRNA^{Scα6,α7} also homozygosed the *Hsα6* and *Hsα7* loci, generating a quintuple-humanized strain (CFU_O/CFU_E =∼1). However, post sporulation, the mix plated on *MATα* selection failed to produce any viable humanized haploid strains. (G) Triple-humanized *Hs-α1α2α3* [haploid *MATa*, also harboring a triple-sgRNA CRISPR reagent (pCas9-sgRNA^{Scα1,α2,α3})] and double-humanized *Hs-α4α5* (haploid *MATα*) strains were mated generating a heterozygous diploid strain for 5 loci. The resulting diploid strain simultaneously combined three yeast to humanized *Hsα1, Hsα2, and Hsα3* loci while harboring two yeast-human heterozygous loci (*α4* and *α5)* (CFU_O/CFU_E =∼1). The subsequent MERGE^0 strategy (using pCas9-sgRNA^{Scα4}) generated a viable quadruple-humanized strain (*Hs-α1α2α3α4*) with heterozygous yeast-human *α5* locus (CFU_O/CFU_E =∼1). The locus-specific PCR confirmed the homozygosity for most of the colonies. However, the subsequent CRISPR selection (using pCas9-sgRNA^{Scα5}) did not produce a viable strain (CFU_O/CFU_E =∼0), suggesting an incompatible quintuple-humanized (*Hs-α1α2α3α4α5*) genotype. (H) Using the quintuple-humanized diploid strain *Hs-α1α2α3α4α6* harboring a yeast-human heterozygous *α7* locus as a background, the transformation of pCas9-sgRNA^{Scα7} yielded a viable sextuple-humanized (Hs-*α1α2α3α4α6α7*) genotype (MERGE^0, CFU_O/CFU_E =∼1) after incubation for 8 days. The genotype of the viable colonies confirmed the homozygosity and humanization of the *α7* locus.

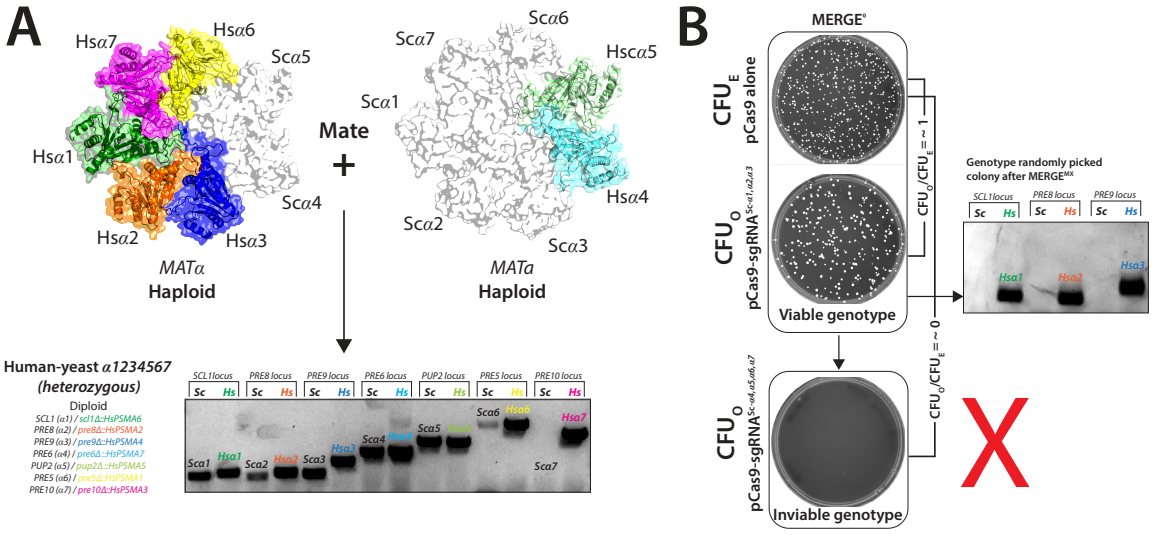


Figure S2.13. MERGE enables the verification of the entire human α-proteasome core in yeast.

(A) Mating quintuple-humanized haploid *Hs-α1α2α3α6α7* (*MATα*) and double-humanized *Hs-α4α5* (*MATa*) strains generated a diploid strain harboring yeast-human heterozygous alleles for 7 *α*-proteasome loci. Locus-specific PCR confirmed the heterozygous nature of the loci. (B) The resulting diploid strain was transformed with a triple-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\alpha1,\alpha2,,\alpha3}$), resulting in a viable genotype homozygous for triple-humanized Hs-*α1α2α3* loci (CFU_O/CFU_E =∼1) with yeast-human heterozygous alleles for *α4, α5, α6,* and *α7* loci. The subsequent transformation of a quadruple-sgRNA CRISPR reagent (pCas9-sgRNA$^{Sc\alpha4,\alpha5,\alpha6,\alpha7}$) in the diploid strain was lethal, suggesting an incompatible genotype (CFU_O/CFU_E =∼0).

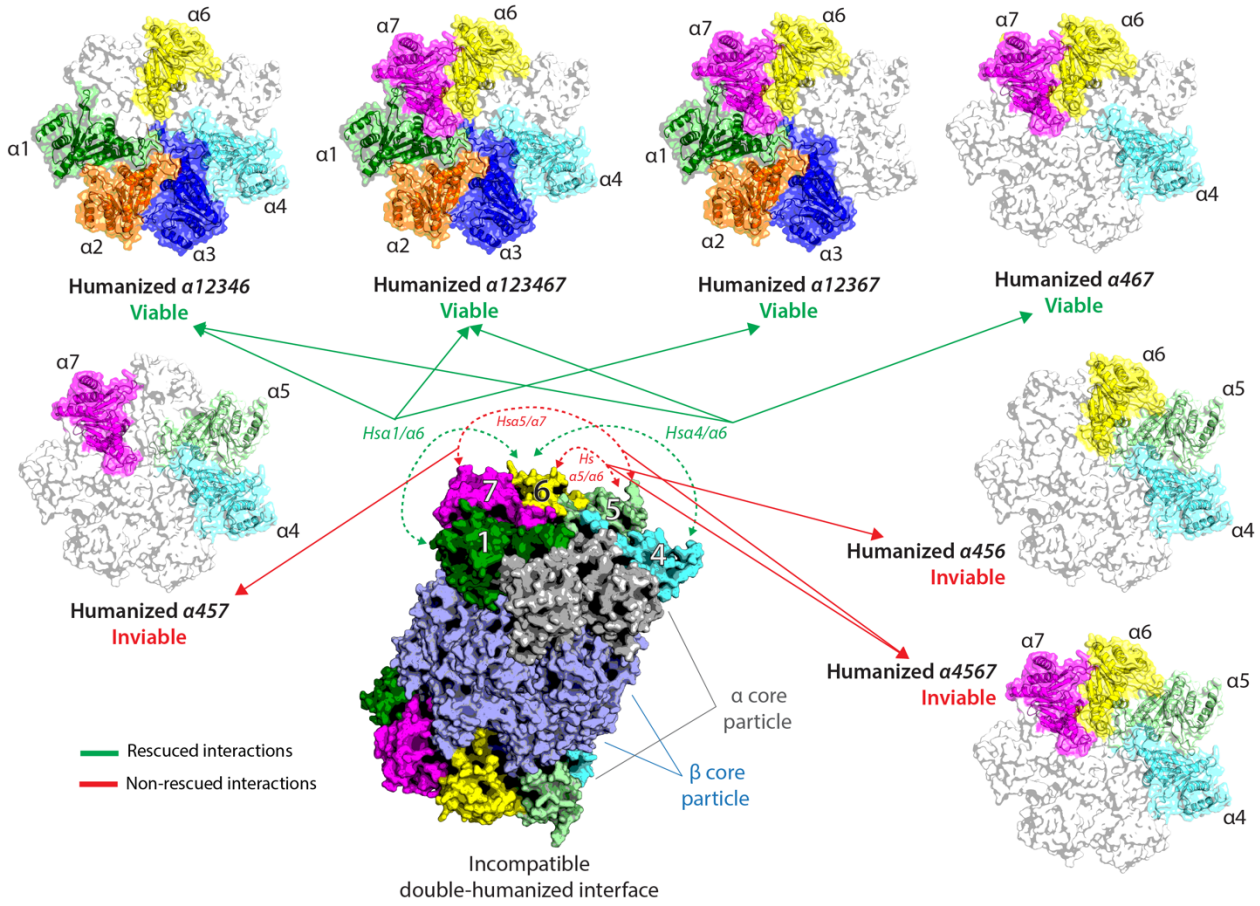


Figure S2.14. Inviable double-humanized genotypes are viable when neighboring interactions are restored.

Inviable double-humanized genotypes, *Hsα1/α6* and *Hsα4/α6,* become viable when the neighboring interactions are restored in higher-order humanized strains (indicated as green lines). However, incompatible paired genotypes comprising *Hsα5/α6* and *Hsα5/α7* could not be rescued (marked as red lines). Proteasome core structures were generated using Pymol and PDB-1RYP. Colored structures show humanized *α* yeast subunits.

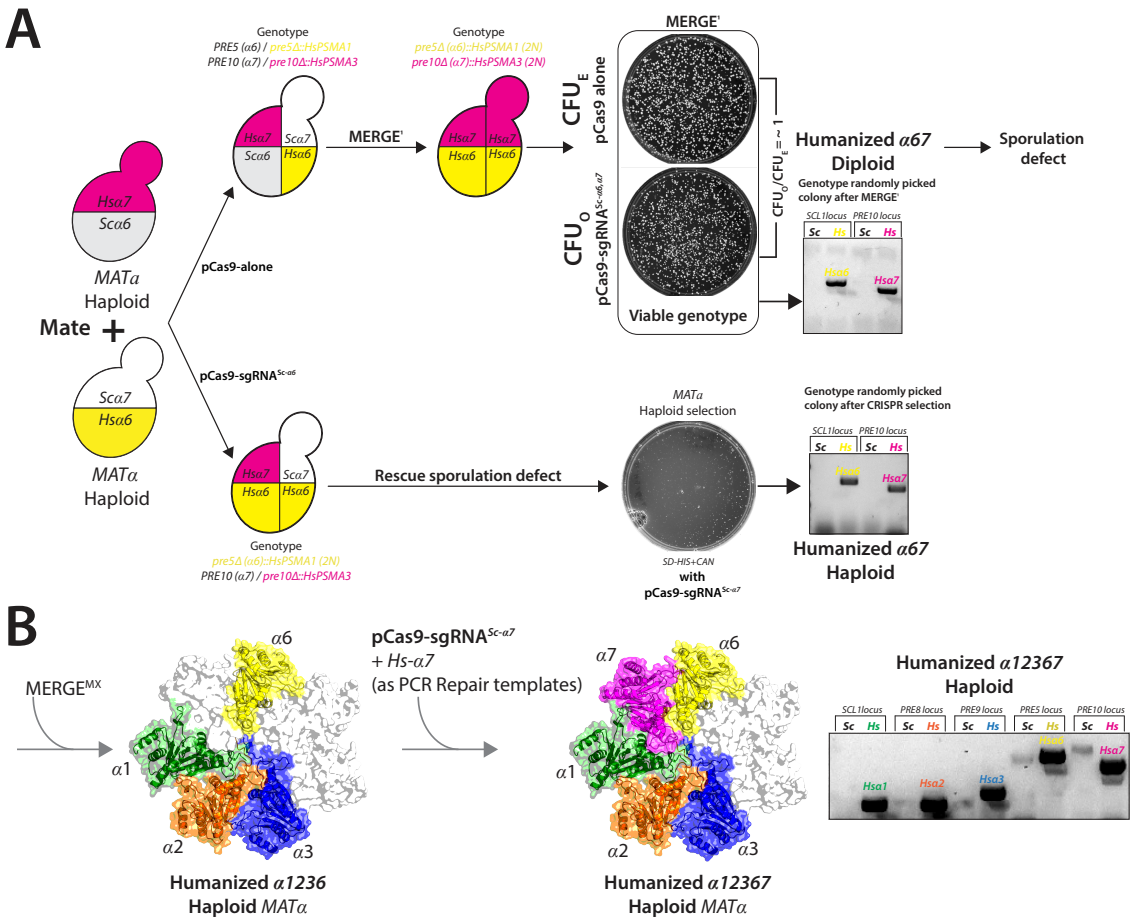


Figure S2.15. Strategies to generate viable humanized haploid yeast strains when the corresponding diploids manifest sporulation defects.

(A) Mating single-humanized *Hs-α6* (*MATα*, with pCas9 alone or pCas-sgRNA$^{Sc\text{-}α6}$) and *Hs-α7* (*MATa*) strains generated two diploid strains; one harboring yeast-human heterozygous allele at *α6* locus and another with homozygous humanized *α6* locus. MERGE^{0}(using strategy pCas-sgRNA$^{Sc\text{-}α7}$) converted the heterozygous yeast-human *α7* locus to a homozygous human allele at ~100% efficiency (CFU_O/CFU_E =~1). Locus-specific PCR confirmed the humanized homozygous alleles. However, the diploid strains cannot sporulate. However, keeping the *α7* locus as heterozygous rescues the sporulation defect. The haploid-specific selection (SD-HIS+CAN; *MATa*) allowed haploid yeast to grow, followed by CRISPR plasmid selection (using pCas-sgRNA$^{Sc\text{-}α7}$) or simply genotyping the haploids to identify homozygous double-humanized Hs-*α6α7* genotype. (B) The diploid quintuple-humanized Hs-*α1α2α3α6α7* strain shows a sporulation defect as in Figure S12F. However, using a sequential CRISPR strategy, a haploid quadruple-humanized Hs-*α1α2α3α6* strain enables the generation of haploid quintuple-humanized Hs-*α1α2α3α6α7* strain to be used for subsequent MERGE experiments.

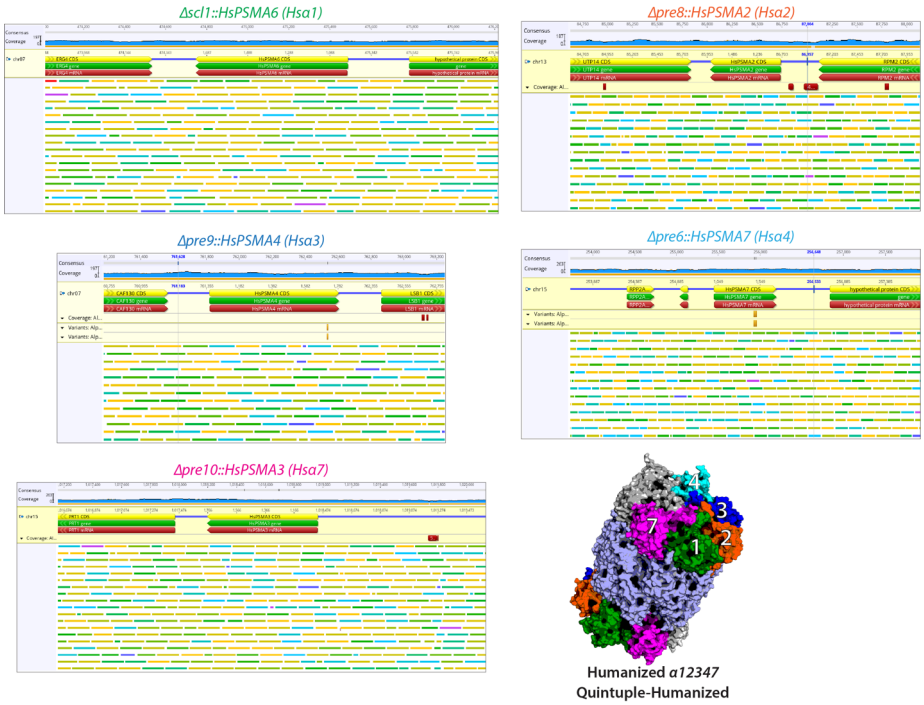


Figure S2.16. Whole-genome sequencing (WGS) of quintupled-humanized *Hs-α1α2α3α4α7* strain shows humanized loci alignment within the yeast genome.

WGS analysis of quintupled-humanized *Hs-α1α2α3α4α7* shows humanized loci aligned to a reference S288C genome with replaced human coding sequences. The human genes reside on different yeast chromosomes, replacing their corresponding native yeast loci (Chr. 7, 13, and 15). Green lines indicate the read lengths with a 200 base range, yellow is below, and blue is above the range.

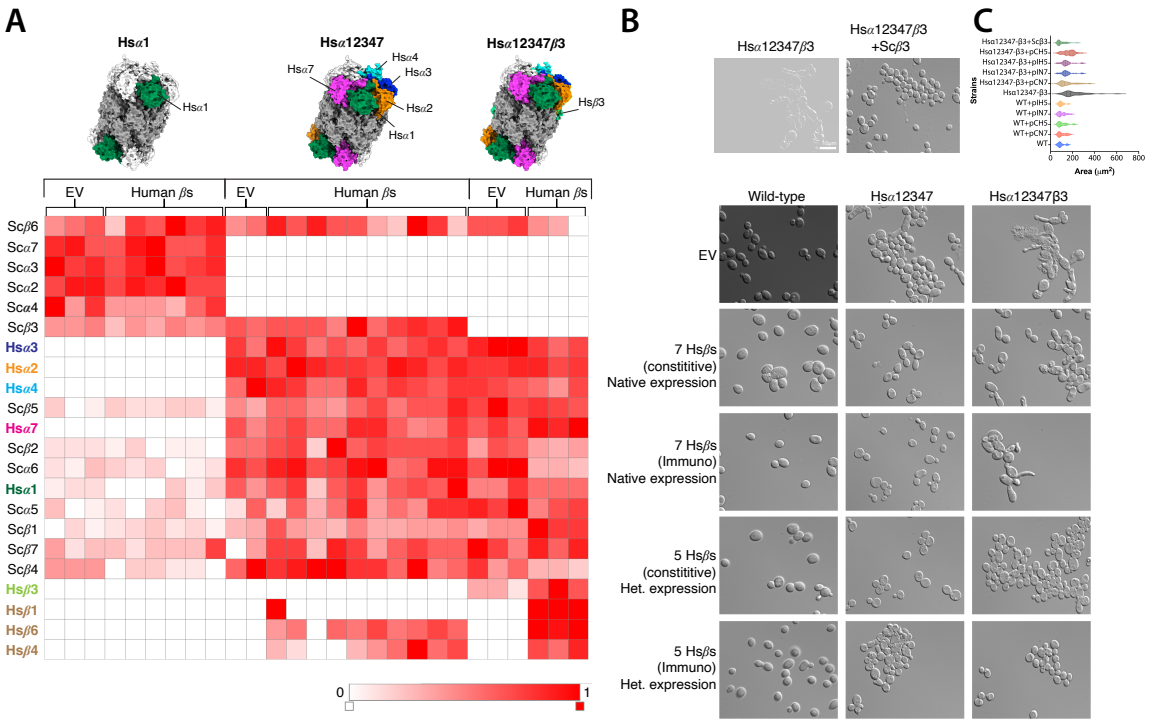


Figure S2.17. Humanized proteasome strains show stable expression of human *β*s.

(A) Heatmap showing abundance measurements for each of the yeast and human 20S proteasome core proteins from single-humanized Hsα1, quintuple-humanized Hsα1α2α3α4α7, and sextuple-humanized Hsα1α2α3α4α7*β*3 strains harboring empty vector (EV) or plasmids expressing human *β*s (heatmap was generated using MORPHEUS software). (B) Brightfield microscope images of humanized yeast cells harboring empty vector (EV), plasmids harboring 7 constitutive- (pCN7) and immuno-*β*s (pIN7) driven by native orthologous yeast gene promoters and terminators, and 5-functionally non-replaceable human constitutive-*β*s (pCH5) and immuno- *β*s (pIH5) driven by strong heterologous yeast promoters and terminators. (C) Humanized strains show varying cell sizes. Cell areas of humanized strains are represented as violin plots on the X axis. Sextuple-humanized Hsα1α2α3α4α7*β*3 strain shows elongated or abnormal cellular morphologies that can be rescued by the expression of the yeast *β*3 (Sc*β*3) or by the coexpression of several human *β*s. Significance comparisons were performed with wild type yeast determined by a standard t-test.

2.6 Tables

Table S1.1-list Oligonucleotides and primers used.

Primer name	**Sequence (5' to 3')**
ScSCL1 (alpha1) _Conf_Fp	CCTCTTATAATGGTAATCGCGATCC
ScSCL1(alpha1)_Conf_Rp	ACAATCTTTACCTCTGACCG
HsPSMA6 (alpha1)_Conf_Rp	CTACTTGGTAGAGCCGACCC
ScPRE8 (alpha2)_Conf_FP	CCTTTCGACGTGAAAAGCGATAATAGC
ScPRE8 (alpha2)_Conf_RP	CTTCTTTTCTGTGGCAATTACTA
HsPSMA2 (alpha2)_Conf_Rp	ACTTCGCTCATCATACAGAA
*ScPRE*9 (alpha3)_Conf_Fp	TTCGTTACCCTGCAGGTGGC
ScPre9 (alpha3)_Conf_Rp	GCGTTCTGCTGCAAGAACAATC
*HsPSMA*4 (alpha3)_conf_Rp	GCCTTAGTTCATTAGTCAGAACA
ScPre6 (alpha4)_conf_Fp-1	GTAAGAATATCATCCGGGCAACGG
Sc*P*re6 (alpha4)_conf_Rp-1	GCGTTGTTCGATAAACCGGG
ScPre6 (alpha4)_conf_Fp-2	AGACCTTCTTTCGCAGCCCAAT
S*cPre6* (alpha4)_conf_Rp-2	ACACCAGCGACATAACGTGT
*HsPSMA7*_conf_Rp	CGCCCATTGCTCTGCGTATA
ScPUP2(alpha5)_conf_FP-1	CAGGGTTATCATAAACCAAGACC
ScPUP2(alpha5)_conf_FP-2	GCGTGCAAACATGAATGTAGC
ScPUP2(alpha5)_conf_RP-1	CACACCTAGTACAACACCTTCTTT
ScPUP2(alpha5)_conf_RP-1	ATGGAAAAGTTGATAACCATC
HsPSMA5-conf_RP	GTCCATATGAAACAGCTGGGG
ScPRE5(alpha6)_conf_FP_1	GAACATACCCTATTATAGAGTGGC
ScPRE5(alpha6)_conf_FP_1	ATCAGTACTAATTTCTATTAGCTGC
ScPRE5(alpha6)_conf_RP_1	TTTCAACGCTACTAAAACAGCAT
ScPRE5(alpha6)_conf_RP_1	CGTATCCTTACCCACGATGGCGAT

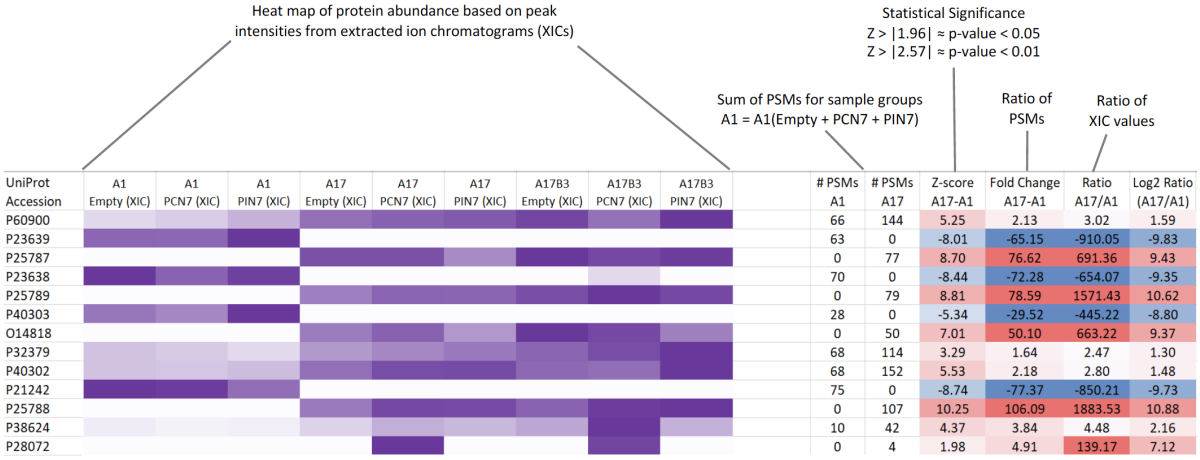
ScPre10(alpha7)_conf_Fp	TGGTAGTGCAGTTGTAAAGCT
ScPre10(alpha7)_conf_Rp	CACTGCAAAGACTACACCGTCGT
HsPSMA3(alpha7)_conf_Rp	GTCTTCCATTAGTTAATTCACCA
*511b*_conf_FP	GGAAATAGACTTTTTTTACACAGCAACGG
*511b*_conf_RP	GGCTGACTGGCATACACTGG
*CrtE*_conf_RP	AGGATGTTCGCGTAATCCAT
*Userx*_conf_FP	CGTGTACCCACAATAGCAGCTACC
*Userx*_conf_RP	GTTGCTTTTCTGTTCCCCTTGAG
CrtI-conf_RP	GACCTGGAAACCTTCTTTAGC
*Fgf20*_conf_FP	CACACGTCGGCACTTACTATCC
*Fgf20*_conf_RP	CTCTGAGGATATAGGAATCTACAAAATG
CrtYB-conf_RP	GGATATGTCCATGCGCCATTT
*SgRNA1-SCL1*Forward primer	GACTTT**CATCACTATCTTTTCCCCCG**
*SgRNA1-SCL1*Reverse primer	AAAC**CGGGGGAAAAGATAGTGATG**AA
*SgRNA2-SCL1*Forward primer	GACTTT**ATAAACTCACTAGCGGTCAG**
SgRNA2-SCL1 Reverse primer	AAAC**CTGACCGCTAGTGAGTTTAT**AA
SgRNA-PRE8 Forward primer	GACTTT**TTGGGTATAAAAGCTACGAA**
SgRNA-PRE8 Reverse primer	AAAC**TTCGTAGCTTTTATACCCAA**AA
*SgRNA1-PRE9*Forward primer	GACTTT**ACAACAATTTTCTCCCCTGA**
*SgRNA1-PRE9*Reverse primer	AAAC**TCAGGGGAGAAAATTGTTGT**AA
*SgRNA2-PRE9*Forward primer	GACTTT**TTGGGATTATGGCATCTGAT**
*SgRNA2-P*RE9 Reverse primer	AAAC**ATCAGATGCCATAATCCCAA**AA
*SgRNA1-PRE6*Forward primer	GACTTT**AGATGGACACATTTTCCAAG**
SgRNA1-PRE6 Reverse primer	AAAC**CTTGGAAAATGTGTCCATCT**AA
*SgRNA2-PRE6*Forward primer	GACTTT**AGTGGAGTACGCCCTGGAGG**
SgRNA2-PRE6 Reverse primer	AAAC**CCTCCAGGGCGTACTCCACTAA**
*SgRNA1-PUP2*Forward primer	GACTTT**ACTAGAAGTGAATATGATCG**

SgRNA1-PUP2 Reverse primer	AAAC**CGATCATATTCACTTCTAGT**AA
*SgRNA2-PUP2*Forward primer	GACTTT**GCACATTTTCCCCAGAAGGG**
SgRNA2-PUP2 Reverse primer	AAAC**CCCTTCTGGGGAAAATGTGC**AA
*SgRNA1-PRE5*Forward primer	GACTTT**GTTCAGGAACAATTACGACG**
SgRNA1-PRE5 Reverse primer	AAAC**CGTCGTAATTGTTCCTGAAC**AA
*SgRNA2-PRE5*Forward primer	GACTTT**GTTCCAAGTGGAATACGCCT**
SgRNA2-PRE5 Reverse primer	AAAC**AGGCGTATTCCACTTGGAAC**AA
*SgRNA1-PRE10*Forward primer	GACTTT**ATCGGTATAAAGTGTAACGA**
*SgRNA1-PRE10*Reverse primer	AAAC**TCGTTACACTTTATACCGAT**AA
*SgRNA2-PRE1*0Forward primer	GACTTT**CAAGTCGTAGACCGTCACAT**
*SgRNA2-PRE10*Reverse primer	AAAC**ATGTGACGGTCTACGACTTG**AA
sgRNA-511b Forward primer	GACTTT**CAGTGTATGCCAGTCAGCCA**
sgRNA-511b Reverse primer	AAAC**TGGCTGACTGGCATACACTG**AA
sgRNA-fgf20 Forward primer	GACTTT**GTTAGAGCTGTTACAAGTTA**
sgRNA-fgf20 Reverse primer	AAAC**TAACTTGTAACAGCTCTAAC**AA
sgRNA-userx Forward primer	GACTTT**GTAGCTACAAGAACATATGG**
sgRNA-userx Reverse primer	AAAC**CCATATGTTCTTGTAGCTAC**AA
sgRNA-ADE2 Forward primer	GACTT**TAATTGTAGAGACTATCCACA**
sgRNA-ADE2 Reverse primer	AAAC**TGTGGATAGTCTCTACAATT**AA
sgRNA-GFP(S65T) Forward primer	GACTTT**GGTGAAGGTGATGCAACATA**
sgRNA-GFP(S65T) Reverse primer	AAAC**TATGTTGCATCACCTTCACC**AA
sgRNA-PUP3-Forward primer	GACTTTA**TTGCCTGTGATTTGCGTCT**
sgRNA-PUP3-Reverse primer	AAAC**AGACGCAAATCACAGGCAAT**AA
sgRNA-kanMX Forward Primer	GACTTT**TGTTTTGCCGGGGATCGCAG**
sgRNA-kanMX Reverse Primer	AAAC**CTGCGATCCCCGGCAAAACA**AA
*PSMA6*_Repair_FP	**CGTAAGGATAGAGTAAGTTAAATGGCTAACTCATTATAATCTTCATGCTAAA TCATATAAGGGCAGAGACGAAGCAAAGCGAAAAAAACATATTACAATC**ATGT CCCGTGGTTCCAGCG
*PSMA6*_Repair_RP	**TCTACTTTTTCCTATAAAATAAATCCTAAATGAAACGTGTCCATACATATATT GAGTGTTGACGCGTGTGATTTCACATTATGTTGTGGCAGGAAGTATG**TTAGT CTCTCTCTGCTAGAGCAAC

*PSMA2*_Repair_FP	**CCCCAATAAGCTGAGAGTGGAATAGGTGAAGCATTTTATATTTTTAGTTTCA ATTAGTAGTAAGCAACCATAAGACACCAATCAACACAGTTCTATAATT**ATGG CGGAGCGCGGGTAC
*PSMA2*_Repair_RP	**TAGAAAAAAATCATTTAAAGGTTATGTAAGTTATATACACAATTGGCTTTCT TTTGGATAAGTTGAGTGAGATGGGTGATTGGCGGGGATAATTATAACA**TTAT GCTATGGCAGCCAAGTAATC
*PSMA4*_Repair_FP	**ACATTGGCAGAGCGAAGAGAACAGACTGCTTTCTATAAAAAGTTTTCGATCA GTCTCTATTTTAATAATTGATTATTGGATATAGTTAGTAGTGTTAAAC**ATGTC TCGAAGATATGACTCCAG**G**
*PSMA4*_Repair_RP	**GGGAATCTTAGCGATTTTTTTCTTTCACTTTTAATTTTAGAGATATGTTTCTA TGCGTACATATTTATATAAGCATGAAGTCAAACAATACTTTCCAACC**CTATTT ATCCTTTTCTTTCTGTTCTTTTTC
PSMA7-Repair_FP	**GGAAGGGAAAGAAAAGACATCATATATAGGTGGTAGCAAAACCTTCAAGTT AGTGCTCAGCAGTTCATCAGGATCAGTTGCATAAACACAATTACACAAA**ATG AGCTACGACCGCGCC
PSMA7-Repair_RP	**AAACATATATGAGCTAACTATACCGTGCTATGTTATATTTTGCATTTTTATTA TTGCTGTTATTTTATATAGGTTTTATGCCCAATATATATCGCCGTTTT**CATGA TGCTTTCTTTTGTTTCTTCTTTTCG
*PSMA5*_Repair_FP	**GCAAAAGAGTAACCTAGGAAATTGGACAAGAGAGCGCTATAACTATAGAAT ACAAGTCAATAAGAACATAAATTCCAATTGTCTAATATAACTAGAAAGG**ATG TTTCTTACCCGGTCTGAG
*PSMA5*_Repair_RP	**TGTGATCAATTGAGGTTATTTTACTACTTTTCCTTTCATTTTTGTAAGGTTTT CTTTCTTTGTTAGTGTGACGTTGGTATTTACCTTTATGTAACTATAT**TTAAAT GTCCTTGATAACCTCTTCAAG
*PSMA1*_Repair_FP	**ACCAAGACGCCACTACATTTAGGTAGACGTGGTGATCAATTGAAGTTCAAG AGAAGTGGGTAAGAATACATTGAACTAAGCCTTCATATATTTACAAAAC**ATG TTTCGAAATCAGTATGACAATG
*PSMA1*_Repair_RP	**ATTATTATCTTTATATGTAAAAAAACCCCGACCGTGGTGAAAGGCTTCTTTG GTCTTGCTTTGACAATTGCTTAGTGGAAACGATACTTTGTGTCCAGAC**TTAA TGTTCCATTGGTTCATCAGCC
*PSMA3*_Repair_FP	**ATAGTGGCGGTAGAGGGTTTATTTGGTAGTGCAGTTGTAAAGCTTTCAAAC ACATTCGAGCGTCGCATCATCAAATTAACAAAAGCATAACTCTTCAGCA**ATG AGCTCAATCGGCACTGG
*PSMA3*_Repair_RP	**ATGGAAGGTACTAAAATGTATTTTTTGTTTCGCGGCGTTACACGTGAGTTCA TATTATTTCAACTCTTTGGTTCTTCTTAACGTATTATCAGAATGTCACTTACA TA**TTATCATCATCTGATTCATCTTC
crtE-Repair_FP	**CCCAACAATACAATAGCGGGAAGAATGCACTTTCTCGTTCCATAAAGGGTCT CTTTCACCTATACGGTTGGTACAGATTTCC**TAGCGGCGCCGGGAAATC
*crt*E-Repair_RP	**GCTTCACTTTTTTCTCTGTCAAGGGACAGTTGCATACAAACGCAGTTCAAAA CGAAGAAAATAGAAGCAAACGACGTAATGC**GCCGCGAGTCACTACTAA
*crtYB*_Repair-FP	**GAAAGAAAGAAAGAAAAACTAACACATTAATGTAGTTTTAAAATTTCAAATC CGAACAACAGAGCATAGGGTTTCGCAAA**CTTGAATCGGCTTTAGGATC
*crtYB*_Repair-RP	CGAGCTTGTAGCACAATAATACCGTGTAGAGTTCTGTATTGTTCTTCTTAGTGCTT GTATATGCTCATCCCGACCTTCCATTAGCCGCAATCGTATCTGA
*userX*_Repair_FP	**GGAGTTGTATCGATGTGTAGTAACTGGTAGTAGTCTAGTAATGAGTTGTTCT CATGTTAGTGAATATTCTGTATGCTTCC**GAGTCTACGTTACACCTGAA
*userX*_Repair_FP	**ATCAAATGAAAGTAATAATCTTTCAGCTTAGACTTATTTTTAGTTGAAACAC TTGAATAAGAGTGCAGACACTCAATCGTGG**TGGGGACACCATCCCGCC
sgRNA_DKO_TU_ConLS (FpS)	ATGCCGTCTCACTGAATGTGCTTCAGTATTACATTTTTTGCCTTC
sgRNA_DKO_TU_ConL1 (Fp1)	ATGCCGTCTCACCAAATGTGCTTCAGTATTACATTTTTTGCCTTC
sgRNA_DKO_TU_ConL2 (Fp2)	ATGCCGTCTCAGATGATGTGCTTCAGTATTACATTTTTTGCCTTC
sgRNA_DKO_TU_ConL3 (Fp3)	ATGCCGTCTCAGTTCATGTGCTTCAGTATTACATTTTTTGCCTTC
sgRNA_DKO_TU_ConL4 (Fp4)	ATGCCGTCTCAGGTAATGTGCTTCAGTATTACATTTTTTGCCTTC
sgRNA_DKO_TU_ConL5 (Fp5)	ATGCCGTCTCAAAGTATGTGCTTCAGTATTACATTTTTTGCCTTC
sgRNA_DKO_TU_ConL6 (Fp6)	ATGCCGTCTCAAACGATGTGCTTCAGTATTACATTTTTTGCCTTC

sgRNA_DKO_TU_ConR1 (Rp1)	ATGCCGTCTCATTGGTATCCACTAGACAGAAGTTTGCGTTC
sgRNA_DKO_TU_ConR2 (Rp2)	ATGCCGTCTCACATCTATCCACTAGACAGAAGTTTGCGTTC
sgRNA_DKO_TU_ConR3 (Rp3)	ATGCCGTCTCAGAACTATCCACTAGACAGAAGTTTGCGTTC
sgRNA_DKO_TU_ConR4 (Rp4)	ATGCCGTCTCATACCTATCCACTAGACAGAAGTTTGCGTTC
sgRNA_DKO_TU_ConR5 (Rp5)	ATGCCGTCTCAACTTTATCCACTAGACAGAAGTTTGCGTTC
sgRNA_DKO_TU_ConR6 (Rp6)	ATGCCGTCTCACGTTTATCCACTAGACAGAAGTTTGCGTTC
sgRNA_DKO_TU_ConRE (RpE)	ATGCCGTCTCATGCTTATCCACTAGACAGAAGTTTGCGTTC

Table S2.2 Protein abundance heat map



UniProt Accession	A1 Empty (XIC)	A1 PCN7 (XIC)	A1 PIN7 (XIC)	A17 Empty (XIC)	A17 PCN7 (XIC)	A17 PIN7 (XIC)	A17B3 Empty (XIC)	A17B3 PCN7 (XIC)	A17B3 PIN7 (XIC)	# PSMs A1	# PSMs A17	Z-score A17-A1	Fold Change A17-A1	Ratio A17/A1	Log2 Ratio (A17/A1)
P60900										66	144	5.25	2.13	3.02	1.59
P23639										63	0	-8.01	-65.15	-910.05	-9.83
P25787										0	77	8.70	76.62	691.36	9.43
P23638										70	0	-8.44	-72.28	-654.07	-9.35
P25789										0	79	8.81	78.59	1571.43	10.62
P40303										28	0	-5.34	-29.52	-445.22	-8.80
O14818										0	50	7.01	50.10	663.22	9.37
P32379										68	114	3.29	1.64	2.47	1.30
P40302										68	152	5.53	2.18	2.80	1.48
P21242										75	0	-8.74	-77.37	-850.21	-9.73
P25788										0	107	10.25	106.09	1883.53	10.88
P38624										10	42	4.37	3.84	4.48	2.16
P28072										0	4	1.98	4.91	139.17	7.12

A1	Hs-α1				
A17	Hs-α12347				
A17B3	Hs-α12347$_{ß3}$				
EV	Empty vector				
pCN7	Plasmid expression human constitutive betas under native regulation				

Chapter 3

Species-specific protein-protein interactions govern the humanization of the 20S proteasome in yeast

Sarmin Sultana, Mudabir Abdullah, Jianhui Li, Mark Hochstrasser, Aashiq H. Kachroo

bioRxiv. 2022. Doi: 10.1101/2022.06.20.496808

Abstract

Yeast and humans share thousands of genes despite a billion years of evolutionary divergence. While many human genes can functionally replace their yeast counterparts, nearly half of the tested shared genes cannot. For example, most yeast proteasome subunits are humanizable, except subunits comprising the *β*-ring core, including *β2* (HsPSMB7). We developed a high-throughput pipeline to humanize yeast proteasomes by generating a large library of Hs*β*2 mutants and screening them for complementation of yeast *β*2 (ScPUP1). Variants capable of replacing ScPUP1 included (1) those impacting local protein-protein interactions (PPIs), with most affecting interactions between the *β*2 C-terminal tail and the adjacent *β*3 subunit, and (2) those affecting *β*2 proteolytic activity. Exchanging the full-length tail of human *β*2 with that of ScPUP1 enabled complementation. Moreover, wild-type human *β*2 replaced yeast *β*2 if the adjacent human *β*3 subunit was also provided. Unexpectedly, yeast proteasomes bearing a catalytically inactive HsPSMB7-T44A variant blocking precursor autoprocessing were viable, suggesting an intact propeptide stabilizes late assembly intermediates. Our data reveal roles for specific PPIs governing functional replaceability across vast evolutionary distances.

3.1 Introduction

Despite the divergence of humans and yeast from a common ancestor over a billion years ago, the yeast *Saccharomyces cerevisiae* still shares nearly a third of its proteome with humans (Remm *et al.*, 2001; Sonnhammer *et al.* , 2015). Systematic studies have discovered many conserved human genes that complement a lethal growth defect conferred by the loss of the corresponding shared yeast gene, indicating functional conservation (Kachroo *et al.*, 2015, 2017; Laurent *et al.*, 2020; Garge *et al.*, 2020; Sun *et al.*, 2016; Yang *et al.*, 2017; Hamza *et al.*, 2020, 2015). These studies reveal a striking trend: functional replaceability is not well-explained by sequence similarity between the human and yeast genes. Instead, it is a property of specific protein complexes and pathways referred to as "genetic modularity," wherein some systems are near entirely replaceable, whereas some modules are entirely non-replaceable (Kachroo *et al.*, 2015). The data suggest, for the functionally replaceable set, that the conserved human genes can generally assimilate within the yeast genetic or protein-protein interaction (PPI) networks.

Overall, however, more than 50% of human genes could not replace their yeast orthologs (Kachroo *et al.*, 2015; Laurent *et al.*, 2020). There could be several reasons for this lack of replaceability, such as divergence of important amino acid residues or domains that are required for maintaining function in the yeast orthologs by co-evolution with the rest of the cellular machinery or simply a change of a function. The second mechanism is unlikely for most proteins in deeply conserved complexes and pathways (Dolinski et al. , 2007; Kachroo *et al.*, 2022). For example, nearly all subunits of the yeast proteasome complex are individually replaceable by their human counterparts except for two 'sub-modules', a 5-subunit contiguous subset of the heptameric *β*-ring of the proteasome core, and a pair of interacting proteins (*Rpn3* and *Rpn12*) in the lid subcomplex (Figure S3A) (Kachroo *et al.*, 2015). Surprisingly, yeast and human proteasome core subunits share

only ~50% identical amino acids, even at successfully humanized protein interfaces (Kachroo *et al.*, 2015). The proteasome is a highly conserved molecular machine found in all domains of life, and is primarily responsible for selective protein degradation (Maupin-Furlow, 2011). The eukaryotic 26S proteasome complex comprises two subcomplexes – the 19S regulatory particle (RP) and the 20S core particle (CP) (Figure S3A) (Tomko et al., 2013). The 19S RP binds the 20S CP barrel on one or both ends and is responsible for recognizing ubiquitinated proteins, unfolding them and translocating them into the central CP chamber. The CP, by virtue of its protease active sites within the central chamber, degrades substrates into small peptides. The 20S CP comprises four stacked heptameric rings. Among the four rings, the two outer rings bear seven different but related α subunits, and the two inner rings each have seven distinct *β* subunits. The CP α-rings form gated channels allowing substrate entry to the proteolytic chamber. Three active protease subunits are present in each *β* ring, specifically *β*1, *β*2 and *β*5, which are responsible for catalyzing substrate cleavage (Groll *et al.*, 1999) (Figure S3A). These subunits harbor a catalytic threonine at their N-termini that functions as the attacking nucleophile in peptide-bond hydrolysis, but they are initially synthesized as inactive precursors. Activation occurs by autocatalytic N-terminal propeptide cleavage, which is highly regulated and is initiated only after successful CP assembly, thus only generating the active site threonine residues after they are inside the core (Tanaka, 2009). Among the other *β* subunits, non-catalytic *β*6 and *β*7 subunits are also synthesized with propeptides that are cleaved *in trans* during formation of the mature CP. The remaining *β*3 and *β*4 subunits do not undergo cleavage and retain their primary forms. Notably, none of the human subunits requiring propeptide cleavage can substitute for their yeast orthologs (Figure S3A).

Assembly of the proteasome core occurs through a combination of the self-assembly of subunits guided by the N-terminal and C-terminal regions of specific subunits and external assembly

chaperones (Tomko et al., 2013). Subunits are added sequentially during the assembly process, creating a series of detectable intermediates (Ramos et al., 2008). The incorporation of *β* subunits usually follows α-ring assembly, starting with the *β*2 subunit, at least in mammalian cells (Hirano *et al.*, 2008). A unique C-terminal extension of *β*2 is crucial for viability and wraps around *β*3 and interacts with *β*4, followed by the incorporation of *β*5, *β*6, *β*1 and lastly, *β*7 (Budenholzer *et al.*, 2017) Given the remarkably high conservation of proteasome complex genes, structure, and assembly mechanism between humans and yeast, it is puzzling why the majority of the subunits in the yeast *β* core are not replaceable by their human counterparts (Figure S3A).

Previously, we discovered that a single mutation (S214G) in the human *β*2 subunit PSMB7 (a constitutive $\beta2^{C}$ subunit), which normally cannot substitute for the orthologous yeast *β*2 subunit ScPUP1, conferred functional replaceability (Kachroo *et al.*, 2015) (Figure S3B). However, the basis of this replaceability remained unknown, particularly since this serine residue is conserved in ScPUP1. Given the proximity of Ser214 to the neighboring yeast *β*6 subunit, we hypothesized that the suppressor might restore local PPIs and promote the assembly of the human protein into the yeast CP (Figure S3C). To systematically address the role of suppressor mutations in replaceability and identify incompatible interfaces, we devised a novel screen to obtain replacement-competent human gene variants in a high-throughput manner, providing insights into the likely mechanisms of replaceability of human *β*2 in yeast. We identified multiple replacement-competent variants. Structural modeling of the human mutants indicated at least two modes of suppression: (1) mutations close to the interacting surfaces of neighboring proteins in the complex suggested multiple PPIs critical to *β*-core assembly. Specifically, the mutations in the C-terminal tail extension of human *β*2 suggested the importance of PPIs with the *β*3 subunit for optimal assembly. (2) Mutations that affected the catalytic activity of the human *β*2 protein also enabled assembly into

the yeast proteasome core, presumably through retention of the 43-residue propeptide. To further characterize the role of the C-terminal extension mutations in functional replaceability, we showed that swapping the entire tail from yeast *PUP1* to human *PSMB7* (Tail- Swap1 or TS1) confers functional replaceability in yeast. Biochemical analysis of the human PSMB7-T44A variant revealed a catalytically inactive subunit with no trypsin-like activity. By contrast, variants such as HsPSMB7-S214G and HsPSMB7-TS1 harbored comparable trypsin- like activities as the orthologous yeast protein within the proteasome. Finally, humanization of the Hsβ3 subunit enabled functional complementation by wild-type human β2, generating a doubly humanized Hsβ2β3 strain. Thus, our data show the divergence of local physical interfaces between the human and yeast β core and demonstrate that restoring these interactions enables the humanization of yeast proteasomes.

3.2 Results and Discussion

3.2.1 A novel high-throughput pipeline to screen for yeast-complementing human gene variants.

Our previous screening strategy to identify functionally replaceable human gene suppressors, while successful, was tedious and identified only two suppressors (Kachroo *et al.*, 2015). The technique required manual isolation of colonies followed by tetrad dissection, and identification of suppressors required manual screening of hundreds of yeast colonies. Therefore, we developed a novel pipeline to screen for suppressor plasmids in an automated high-throughput manner (**Figure 3.1A**). We specifically focused on screening for suppressor mutations in the non- complementing human $\beta 2^{C}$ or HsPSMB7, a constitutively expressed proteasome subunit. The method allowed large-scale and error-free screening in a significantly shorter time. Assays were performed in a 96-

well format and haploid-specific mutant isolation using Synthetic Genetic Array (SGA) or Magic Marker (MM) selection to eliminate the need for tetrad dissection (Pan *et al.*, 2004; Kuzmin *et al.*, 2016). A plasmid-dependency assay using 5-FOA selection was used to confirm that complementation was associated with the variant human gene on the *URA3*-based plasmid. Using this strategy, we successfully obtained multiple suppressors in the *HsPSMB7* gene.

The screening pipeline was developed based on several considerations. Wild-type human PSMB7 (*β2*) cannot normally replace the orthologous yeast *β2* gene *ScPUP1* (Kachroo *et al.*, 2015), and PUP1 is required for yeast viability. By contrast, if a human gene (or its variant) successfully replaces the function of the host gene, the strain will be able to grow, serving as a simple readout for functional replacement. The error-prone PCR strategy generated a *HsPSMB7* mutant gene library, with an average of 1-4 mutations per gene, in a *URA3*-marked yeast expression vector (**Figure 3.1A**) To determine if any mutant *PSMB7* alleles can complement the lethality caused by the deleted yeast gene, we transformed the mutant library into a yeast diploid heterozygous knockout *PUP1*/*pup1*Δ*::kanMX* strain (Pan *et al.*, 2004). The transformation protocol was scaled to obtain several thousand well-separated yeast colonies, each carrying a plasmid with a different *PSMB7* mutant (**Figure 3.1B**)

Several thousand colonies were picked in an automated manner using a QPix 460 robot and spotted in 96-well format on the pre-sporulation GNA medium with G418 selection for the *kanMX*-marked *Pup1*Δ. After sporulation, Magic Marker (MM) medium (-Leu -Arg -His -Ura +CAN) allowed the selection of viable *pup1*Δ*::kanMX* haploid yeast spores (in the presence of G418) harboring plasmids with different human *PSMB7* alleles (Figure 3.1C, bottom panel). As an internal control for sporulation efficiency, we also tested the growth of wild type *PUP1* haploid spores on MM medium (in the absence of G418) (**Figure 3.1C**, top panel). The screen identified 19 colonies

that grew on MM+G418 (see representation images in **Figure 3.1D**) that potentially carried complementing human *PSMB7* variants. The haploid suppressor strains were then tested to determine if suppression was due to the presence of the plasmid-borne mutant human gene. The yeast cells were tested for plasmid dependency based on loss of growth on 5-FOA, which selects against the *URA3* gene. Seven of the 19 suppressors did not survive on 5-FOA medium, suggesting that the human gene variants in these strains were essential for their viability (**Figure 3.1D**).

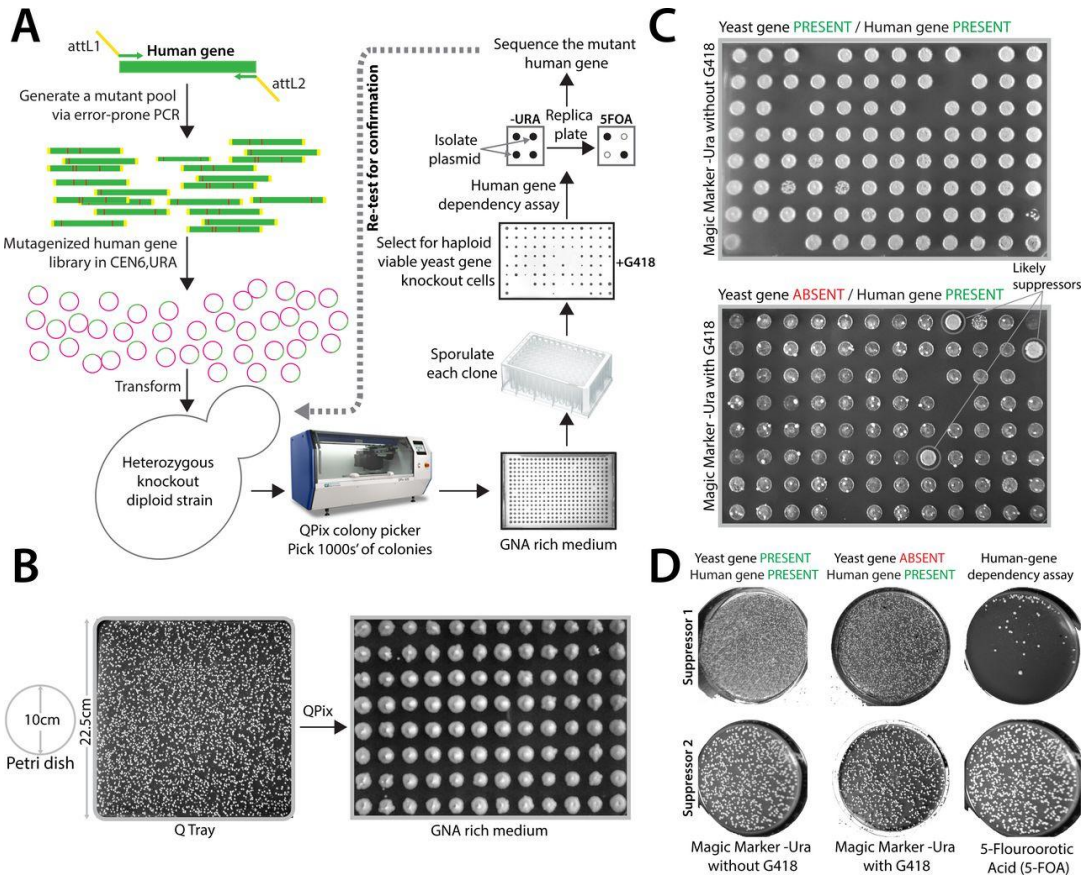


Figure 3.1 High-throughput automated pipeline to screen for functionally replaceable human gene suppressors in yeast.

(A) Workflow showing the generation and screening of a human gene mutant library made by error-prone PCR (0-4 mutations per kbp). The variant pool was cloned into the expression vector (*CEN6*, *URA3*) followed by transformation into the yeast heterozygous diploid knockout *PUP1/pup1Δ::kanMX*strain. (B) Transformation of the mutant library was scaled (4X) and the mixture plated on a QTray. Individual colonies were picked by a QPix 460 colony picking robot (up to 1000 colonies) and spotted on a pre-sporulation GNA-rich media followed by sporulation in a 96-well format. (C) Each sporulation mix was spotted on Magic Marker medium (MM) with (yeast gene absent and human gene present condition) or without G418 (yeast gene present and human gene present condition) to allow the growth of haploid yeast. The likely suppressors appear as spots growing in MM with G418, similar to the MM without G418 condition. (D) Colonies growing on MM with G418 (yeast gene absent and human gene present condition) are further tested by a plasmid-dependency assay using 5-fluoroorotic acid (5-FOA) selection against the *URA3* plasmid. Representative examples showing yeast strains with human gene suppressors that fail to grow on 5-FOA plates, indicating their plasmid dependence (suppressor 1, top panel) or grow on 5-FOA, indicating plasmid independence (suppressor 2, bottom panel). Finally, each suppressor was re-tested to verify functional replaceability and plasmid-dependency, followed by Sanger sequencing to identify mutations in the human gene.

3.2.2 Identification and characterization of complementing human *PSMB7* variants in yeast

To identify the relevant *PSMB7* mutations and quantify *pup1*Δ complementation by the mutant human gene variants, we isolated the plasmid from each strain that had passed the test of plasmid dependency. The extracted plasmids were again tested for functional replaceability in a *pup1*Δ strain. Sanger sequencing of the seven-complementing human *PSMB7* gene variants identified the mutations likely responsible for functional replaceability in yeast. Sequence analysis showed that each plasmid contained 1-4 mutations that resulted in amino acid changes in the human protein. The screen also identified a previously characterized suppressor, Hs*PSMB7*-S214G, further validating the functioning of the pipeline (Figure S3A,). Notably, two independently isolated variants included an active-site Thr44Ala mutation.

All seven original suppressors were confirmed for functional rescue using quantitative growth assays (**Figure 3.2**). We first performed growth assays on solid agar medium. As a control, we used the plasmid-borne cognate yeast gene (pGPD-Sc*PUP1*) as a benchmark for optimal functional replacement (Figure 3.2). The *pup1* knockout strain with pGPD-Sc*PUP1* formed colonies after 2-3 days incubation at 30°C. Expression of wild type HsPSMB7 failed to complement the *PUP1* deletion (**Figure 3.2B**). We tested all seven of the primary Hs*PSMB7* variants by extending the

incubation times up to 6 days, observing a variable level of growth. Hs*PSMB7*-S214G shows a growth profile on solid agar medium identical to wild-type Sc*PUP1*. On the other hand, GPD-driven expression of Hs*PSMB7* variants harboring A70V, T44A-E26K- M67V, or T44A-L116H mutations showed distinctly slower growth in *pup1*Δ cells than did the Hs*PSMB7*-S214G mutant or yeast *PUP1*. Functional replacement by human gene variants with T233R-D96V, S161T-T260I-E263K, or K249R-R32W was less efficient compared to the other variants (**Figure 3.2C**).

To determine which mutations alone or in combination from the original *HsPSMB7* suppressor alleles contributed to the ability to replace yeast *PUP1*, we performed site-directed mutagenesis to generate single-site amino acid substitutions in wild-type *PSMB7*. Each *HsPSMB7* single point mutant was tested again using the previously established pipeline for functional replaceability (**Figure 3.2D**, S3.2A and S3.2B). Eight of 14 single mutants failed to complement *PUP1* loss-of-function (Table 3.1). The remaining six single-site substitutions were identified as efficient suppressors (K249R, S161T, T44A, T233R, A70V and S214G) of the yeast *pup1* deletion (**Figure 3.2C,3.2D**). Quantitative growth assays in liquid cultures revealed growth of the initial *HsPSMB7* variants similar to that of the cognate yeast *PUP1*gene (**Figure 3.2E**). Thus, in the case of these primary suppressors, the additional mutations (L116H, E26K, M67V and N224D) are incidental products of the random mutagenesis and are not required for *PUP1*/*β2* replacement. In contrast, three-point mutants of Hs*PSMB7* (S161T, T233R, and K249R) while able to complement *pup1*Δ, show delayed growth on solid agar and in liquid medium relative to the respective multiply mutant suppressors from which they derive. This suggested accessory roles for the other mutations in the original suppressors (R32W*, D71N*, D96V*, T260I*, E263K*, E263-Stop*; the enhancing ancillary function is indicated by an asterisk) (**Figure 3.2**). It is important to note that our previous plate-based assays failed to identify Hs*PSMB7*- T44A mutant as a

replacement-competent variant (Kachroo *et al.*, 2015). This anomaly resulted from its slower growth rate on synthetic medium agar plates. However, given that we obtained two independent suppressors with the T44A mutation by incubating the Hs*PSMB7*-T44A variant for longer times (additional ~12 hrs.), we conclude that this mutation allows human *β*2 to compensate for the loss of its yeast ortholog.

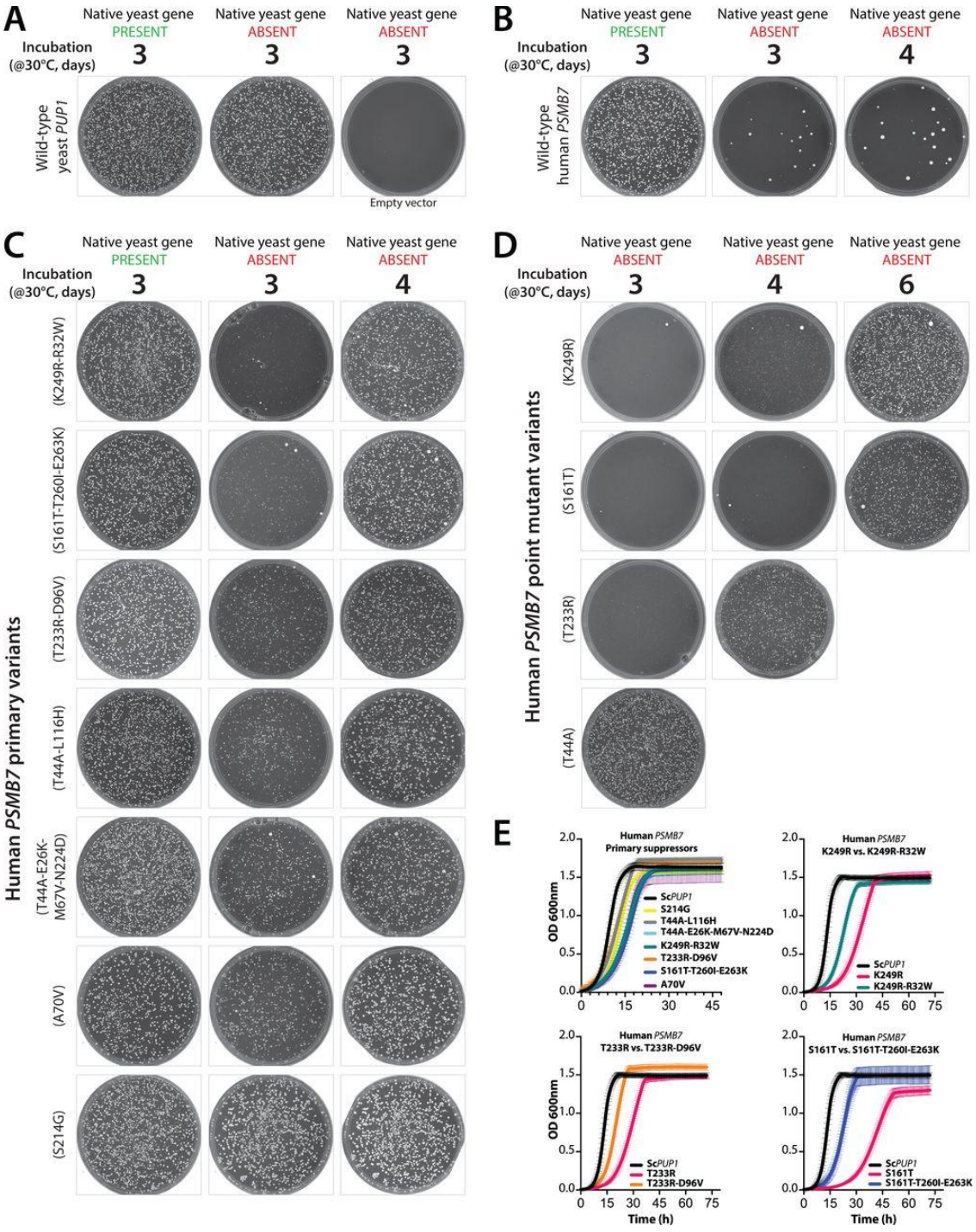


Figure 3.2 Confirmation and quantitative growth analysis of functionally replaceable human gene suppressors in yeast.

(A) Post sporulation selection to grow haploids on Magic Marker medium (MM) with G418 (Yeast gene ABSENT) or without G418 (Yeast gene PRESENT) enables selection for functional replaceability. The expression of yeast PUP1 under the control of the constitutive GPD promoter functionally complements the growth defect of the *pup1Δ::kanMX* strain, whereas the empty vector does not allow growth. (B)The expression of the wild-type human PSMB7 under the control of the GPD promoter does not rescue the lethality of the *PUP1* deletion in MM+G418

(Yeast gene ABSENT). (C) Expression of human *PSMB7* suppressor mutants (K249R-R32W, S161T-T260I- E263K, T233R-D96V, T44A-L116H, T44A-E26K-M67V-N224D, A70V, S214G) on MM-G418 (Yeast gene PRESENT) and MM+G418 (Yeast gene ABSENT) shows variable growth rescue after 3-4 days of incubation at 30°C. The Hs*PSMB7*-S214G variant shows a comparable growth profile to the wild-type yeast *PUP1* gene, whereas the remaining variants show progressively less efficient growth rescue (arrayed from bottom to top). (D) Single-site mutations in human *PSMB7* complement yeast *pup1Δ*. The Hs*PSMB7*-T44A variant alone can complement the lethal growth defect of *pup1* deletion (MM+G418; Yeast gene ABSENT) after 3 days of incubation at 30°C. However, Hs*PSMB7* variants such as T233R, S161T and K249R, while functional in yeast *pup1Δ* cells, show a delayed growth phenotype compared to the primary multiply mutant suppressors from which they derive. (E) Quantitative growth assays confirm the growth pattern of primary human gene suppressors, Hs*PSMB7*-S214G (yellow), Hs*PSMB7*-T44A-L116H (gray), Hs*PSMB7*-T44A-E26K-M67V,N224D (cyan), Hs*PSMB7*-K249R-R32W (green), Hs*PSMB7*- T233R-D96V (orange), Hs*PSMB7*-S161T-T260I-E263K (blue), and Hs*PSMB7*-A70V (purple) display a growth profile similar to the yeast wild type *PUP1* (black) when expressed on a plasmid. By contrast, humanized yeast with Hs*PSMB7* single-site mutants [T233R, S161T, or K249R (pink lines in each graph] show delayed growth compared to the primary suppressors from which they derive, suggesting an accessory role of these other mutations in functional replaceability. The mean of three independent growth curves is plotted with standard deviations.

3.2.3 Structural modeling links mutations at subunit interfaces to functionality of Hs*PSMB7* in yeast

The *HsPSMB7* variants capable of replacing the orthologous *PUP1* gene in yeast are scattered across different regions of the protein with a cluster in the C-terminal tail (**Figure 3.3A**).To explore how these mutated residues might be facilitating complementation, we modeled the HsPSMB7 structure within the yeast proteasome core structure (**Figure 3.3B**).This allowed the classification of the mutations based on their proximity to the neighboring yeast proteasome subunits (**Figure 3.3B)**. A subset of the mutations were classified as ones that likely promote key PPIs with the neighboring yeast proteasome subunits (such as *β*1, *β*3 and *β*6) (**Figure 3.3B**). *PSMB7*-S214G is close to both the catalytic *β2* Thr44 residue (numbering for the precursor form) and the neighboring *β*6 subunit (**Figure 3.3B**). The weakly complementing S161T mutation in Hs*PSMB7* is close to the *β1* subunit interface (**Figure 3.3C**).

The cluster of mutations in the C-terminal tail of Hs*PSMB7* affects interactions with neighboring yeast *β*3 (**Figure 3.3D**). The *β2* C-terminal tail wraps around the outside of the *β*3 subunit in the same ring and also makes contacts with *β*6 in the opposing, C2 symmetry-related *β* ring. Mutations

in the Hs*PSMB7* C-terminal tail (T233R, K249R, T261T*, E263K* and E263- Stop*) impact its interactions with the β3 outer surface, whereas mutations such as A70V, D71N*, and D96V* would affect its binding to β3 at their more interior within-ring interface. The unexpected active-site mutation (T44A) is not close to any interface and is predicted to lead to a catalytically dead β2 subunit that will have an incompletely processed propeptide due to a block in autocleavage.

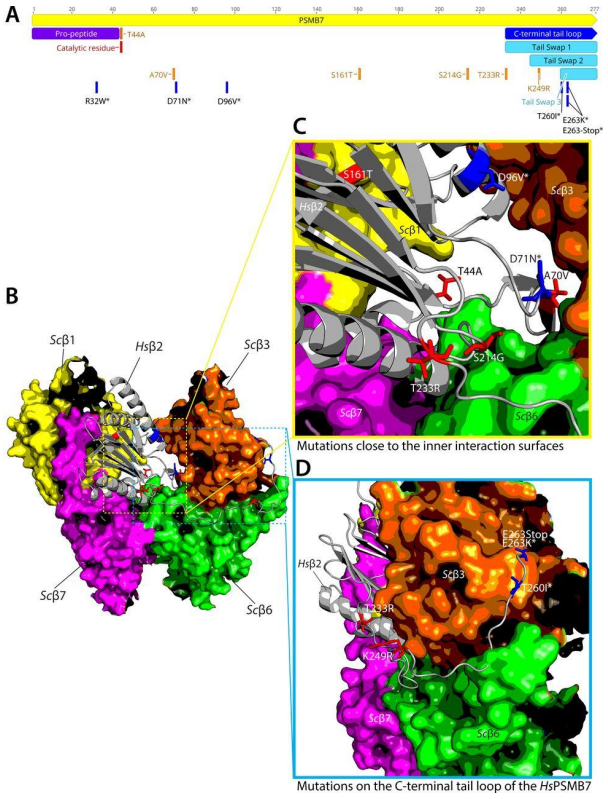


Figure 3.3 Modeling human *PSMB7* variants in the yeast proteasome core suggests a role for specific protein-protein interactions in functional replaceability.

(A) The schematic shows the suppressors in a linear map of Hs*PSMB7*. The C-terminal tail domain is indicated in blue. The mutations with an asterisk (*) do not enable functional replaceability independently but are needed for better replaceability/growth. (B) Human *PSMB7* (gray ribbon; PDB 1IRU) modeled into the yeast proteasome core structure (indicated as colored subunits in yellow- β1, orange- β3, pink- β7 and green- β6; PDB 1RYP). (C) Amino acid changes that contribute to functional replaceability are highlighted (single amino acid substitutions alone in red & accessory residues in blue). An active site substitution T44A in Hs*PSMB7* confers functional replaceability in yeast; however, it does not show proximity to any interacting subunit surface, whereas the remaining mutant residues are close to several subunit interfaces. (D) Replacement- competent Hs*PSMB7* variants, including T233R and K249R and a few accessory substitutions (T260I*, E263K* and E263-Stop), reside on the C-terminal tail of the human protein that wraps around the neighboring yeast β3 protein.

3.2.4 C-terminal tail swaps from yeast to human *β2* show functional replaceability in yeast

The *β2* subunit initiates the assembly of the *β*-ring in the proteasome core, at least in mammalian cells, by recruiting the *β3* subunit (Murata *et al.*, 2009; Tanaka, 2009). As noted above, the unique C-terminal tail of *β2* threads around the *β*3 subunit while also interacting with neighboring *β4, β6*, and *α* subunits (Ramos *et al.*, 2004). Deletion of the yeast *β2* C-terminal tail is lethal and in heterozygous diploids, leads to a pronounced accumulation of assembly intermediates containing the unprocessed *β2* precursor (Ramos *et al.*, 2004). Therefore, the C-terminal tail of yeast *β2* is essential for CP assembly. Sequence alignment of *β2* subunits across diverse species shows that the C-terminal tail has diverged more extensively relative to the rest of the protein (**Figure 3.4A**). Therefore, the incompatibility between the human and yeast *β2* may be a consequence of the divergent interactions of the tail loop, leading to failure of assembly in humanized yeast proteasomes. Indeed, structural modeling of the human *β2* subunit within the yeast proteasome core revealed 5 of 14 Hs*PSMB7* mutations [T233R, K249R, T260I*, E233K* and the previously identified E263- STOP* (Kachroo *et al.*, 2015)] that promoted functional replaceability occurred in the C-terminal tail (**Figure 3.4**).

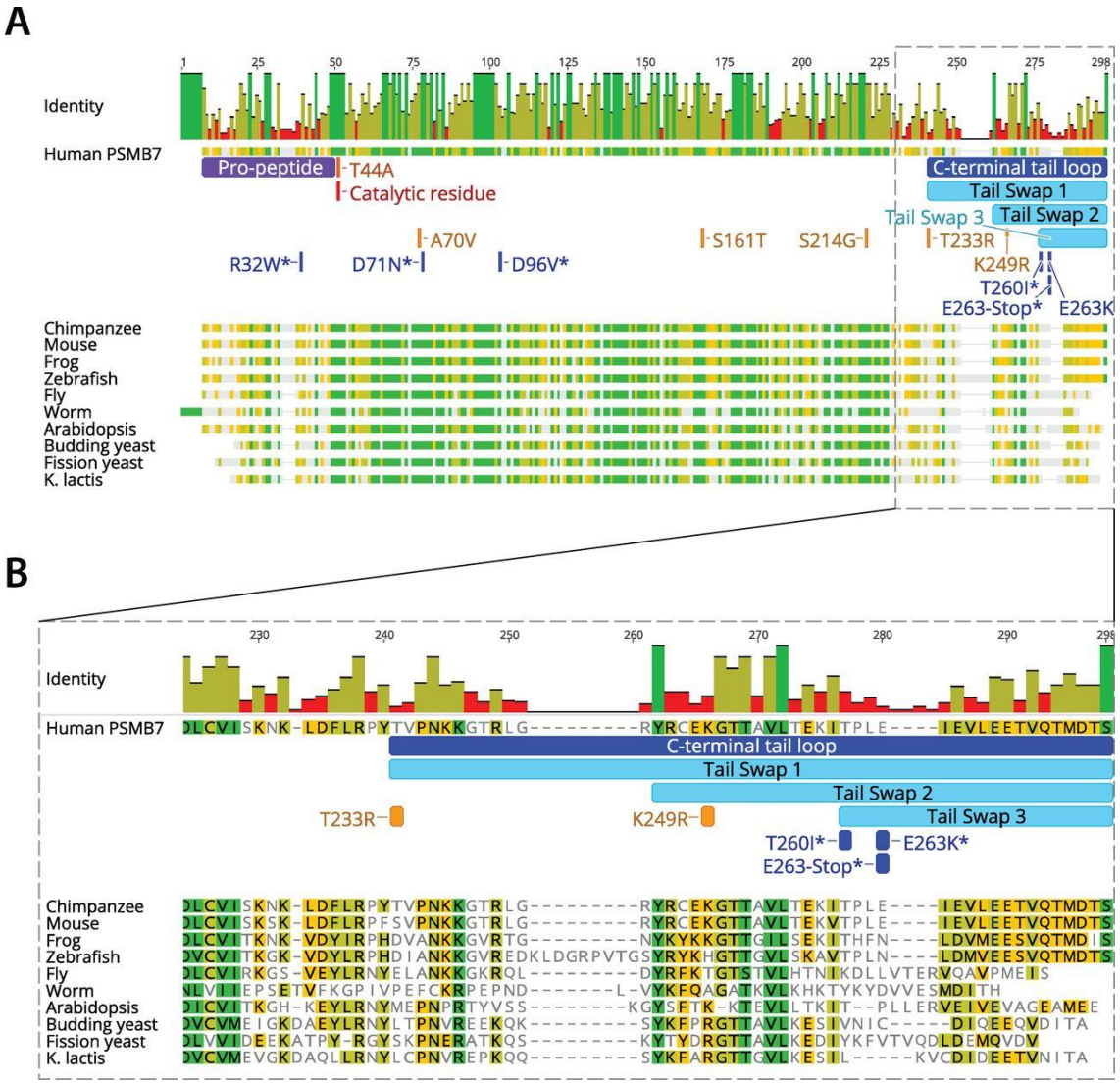


Figure 3. 4. Multiple sequence alignment of *β2* proteins across diverse species highlighting C-terminal tail regions.

(A) The alignment of *β2* protein sequences belonging to diverse organisms shows the divergence of the C-terminal tail of *β2*. C-terminal tails are critical for subunit interaction and sequential proteasome assembly. The identity score is represented as bars with high identity in dark green, medium in light green, and low in red. (B) A magnified view of the aligned C-terminal tail regions highlights the divergent tails. The swaps replacing tail sequences of human *β2* with corresponding yeast sequences are shown in cyan. Tail-Swap1 (TS1) transplants the entire *β2* tail from yeast to human PSMB7, whereas TS2 and TS3 carry C-terminal tail transplants of progressively smaller human tail regions.

Therefore, we asked if swapping segments of the human PSMB7 tail with the corresponding yeast PUP1 tail elements would allow functional replaceability. We engineered three human-yeast hybrid

genes with different lengths of C-terminal tail swaps (**Figure 3.5A**). We chose the lengths of the swapped C-terminal tails based on structural modeling and on the aligned positions of suppressor mutations in the human C-terminal segment (T233R, K249R, T260I* and E263K*). The clones with different lengths of the swapped sequences were designated as Tail-Swap1 (TS1, which included the full yeast PUP1 tail sequence covering the C- terminal tail suppressor mutations in HsPSMB7), TS2 (spanning K249R and downstream mutations) and TS3 (spanning only the accessory mutations T260I* and E263K* and E263-Stop*) (**Figure 3.5A and 3.5B**).These hybrid human-yeast genes were cloned into the yeast expression vector, followed by transformation into *PUP1*/*pup1*Δ*::kanMX* diploid yeast and tested for the functional replaceability. The assay revealed that HsPSMB7-TS1 could replace the *PUP1* ortholog (**Figure 3.5C**). whereas the TS2 and TS3 chimeras failed to complement. The successful rescue was confirmed by plasmid-dependency assays using 5-FOA, which showed that the *pup1Δ* cells cannot survive loss of the Hs*PSMB7*-TS1 plasmid. Quantitative growth assays indicated that cells carrying the human *PSMB7-TS1* chimera grew similarly to the positive control (wild-type yeast *β2*) (Figure 3.5C). Thus, our data demonstrate that the divergence of the C-terminal tails of human and yeast *β2* subunits is likely to have a strong impact on the assembly of the proteasome core.

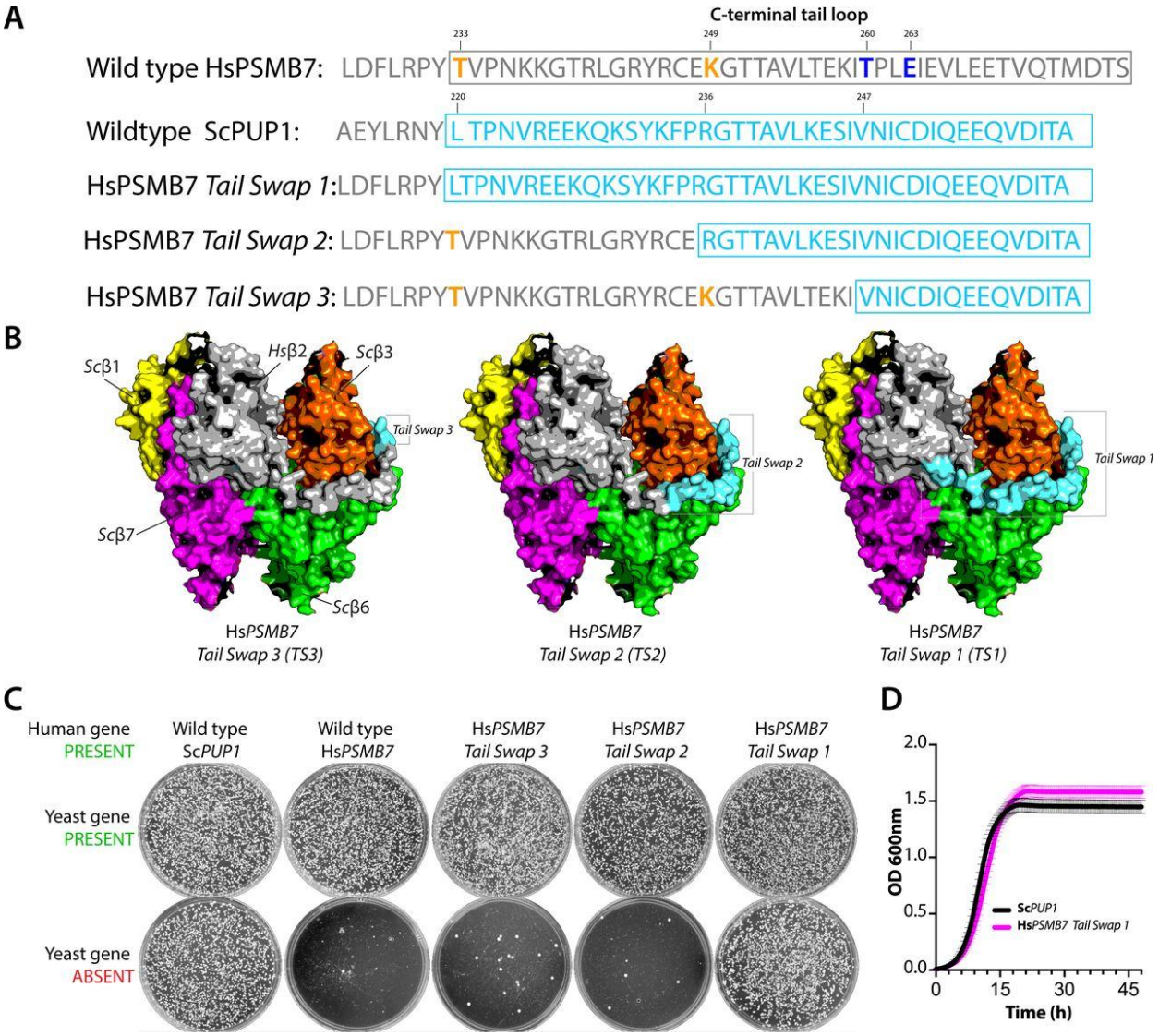


Figure 3.5 Full-length C-terminal tail swap from yeast to wild-type human *β*2 enables functional replaceability in yeast.

(A) The sequences of C-terminal tails of the wild-type human *β*2 (gray), wild-type yeast *β*2 (cyan), chimeric human Tail-Swap1 (TS1, with the entire yeast C-terminal tail), TS2 (intermediate C- terminal tail swap), and TS3 (only the C-terminal-most 15 residues of the yeast tail) are shown. (B) Structure of human *β*2 (gray) (PDB 1IRU) modeled in the yeast *β* core with neighboring yeast *β*3 (orange), yeast *β*1 (yellow), yeast *β*7 (magenta), and yeast *β*6 (green) subunits (PDB 1RYP). The structure shows the C-terminal tail of human *β*2 wrapping around the yeast *β*3 subunit. The C-terminal tail segments of yeast *β*2 that replaced the corresponding human *β*2 segments are shown in cyan. (C) Growth assays were performed on Magic Marker (MM) medium without G418 (yeast gene PRESENT) and MM medium with G418 (yeast gene ABSENT) at 30°C. The data show that plasmid-based expression of yeast PUP1successfully complements the deletion of the native yeast gene copy, whereas the wild-type human *β2* (PSMB7) does not. The full-length C- terminal tail swap from yeast to human *β2* (TS1) showed growth rescue similar to the positive control Sc*PUP1* after 3 days of incubation at 30°C. However, shorter tail swaps did not complement the deletion of *PUP1*. (D) The growth profiles of the human *β*2-TS1 mutant (pink) and the positive control yeast *β*2 (black) are comparable. The mean of 3 independent growth curves is plotted with standard deviation (N=3).

3.2.4 Biochemical characterization of the yeast-complementing human *β*2 variants reveals a catalytically active proteasome core.

Yeast and human *β*2 proteins harbor the catalytic site for a trypsin-like protease activity in the 20S proteasome (Arendt et al., 1997; Harshbarger *et al.*, 2015; Rut et al. , 2016). While several mutations allow human *PSMB7* to complement the lethal growth defect of a yeast *PUP1* deletion, the rescue may occur at the level of proteasome assembly rather than restoration of the *β2* proteolytic activity, which is not essential for viability (Arendt et al. , 1997). The active-site mutant *pup1*-T30A subunit can assemble into the proteasome core. To biochemically characterize human *β2* in hybrid yeast-human proteasomes, we introduced a sequence encoding a 3xFLAG tag at the C-terminal coding sequence of yeast *RPN11* at its native genomic locus using CRISPR-Cas9 and homology-directed recombination (HDR) (Figure S3). *RPN11*, a component of the RP of the 26S proteasome, is not functionally compromised by the 3xFLAG tag (Sakata *et al.*, 2011). The RPN11-3xFLAG strains were confirmed to be stably expressing the tagged protein and showed no growth defects (Figure S3). The humanized *β2* strains were engineered to harbor the RPN11-3xFLAG tag to allow affinity purification of the 26S proteasome (RP-CP).

We tested proteasome activity and assembly in the yeast strains expressing the wild-type yeast *Pup1*, catalytically inactive yeast *pup1*-T30A, human PSMB7-T44A, human PSMB7- S214G, and human PSMB7-TS1 variants of the *β*2 protein (**Figure 3.6**). As expected, trypsin-like activity was abolished in both affinity-purified proteasomes with active-site *β2* mutations, PSMB7- T44A and pup1-T30A. By contrast, trypsin-like activity was observed in the humanized PSMB7- S214G and PSMB7-TS1 proteasomes, as in the positive control, wild-type Pup1 particles (**Figure 3.6A**).

Analysis of the purified humanized proteasome particles by SDS-PAGE followed by anti- PSMB7 immunoblotting revealed that the unprocessed PSMB7-T44A precursor assembled into hybrid 26S

proteasomes, and the particles had no detectable trypsin-like activity; by contrast, the *PSMB7* propeptide had been cleaved in the PSMB7-S214G and PSMB7-TS1 proteasomes (**Figure 3.6A**, PSMB7 blot). Chymotrypsin-like activity, which derives from the yeast *β*5 subunit, was seen in all the purified proteasomes (**Figure 3.6A**). Interestingly, the evaluation of PSMB7 processing activity in whole cell lysates without prior purification showed most of the human *β2* protein still in its precursor form in cells expressing PSMB7-S214G and PSMB7-TS1 (**Figure 3.6B**). This might be due either to a reduced rate of auto processing in the hybrid proteasomes or to the majority of human *β*2 protein being unincorporated or present in immature assembly intermediates. To test this last idea, we evaluated the assembly state of proteasomes in yeast whole cell lysates by nondenaturing-PAGE immunoblot analysis (**Figure 3.6C**). The human *PSMB7* variants were able to assemble into mature yeast proteasomes, but this was accompanied by 20S CP assembly defects and likely concomitant RP base assembly defects. Specifically, an anti-Pre6 (yeast *a*4) blot showed little or no free CP but accumulation of sub-CP species in cells with the humanized *PSMB7* variants (T44A, S214G, and possibly TS1), suggesting limiting amounts of mature CP relative to RP. Anti-Rpt5 blotting revealed an excess of the Rpt4-Rpt5 assembly intermediate (lower band, **Figure 3.6C**). and other particles smaller than full 26S proteasomes. An excess of free RP lid complexes accumulated due to the deficit in fully assembled base intermediates (**Figure 3.6C** middle). Similar proteasome assembly defects, including of the RP, were previously observed in CP assembly chaperone mutants (Kusmierczyk *et al.*, 2008). Assembly defects were less prominent in the PSMB7-TS1 humanized proteasomes; because the C-terminal tail threads along the interface between the two *β* rings, the processing defect observed in this mutant might reflect slower auto processing since this only occurs after two half-proteasomes have associated properly (Chen *et al.,, 1996*). These results suggest that human *PSMB7* variants (S214G and TS1) can replace yeast *Pup1*

catalytic activity and assemble into functional humanized yeast 26S proteasomes. Furthermore, the data confirm that Thr44 of the human PSMB7/β2 is essential for its auto processing and proteasomal trypsin-like activity.

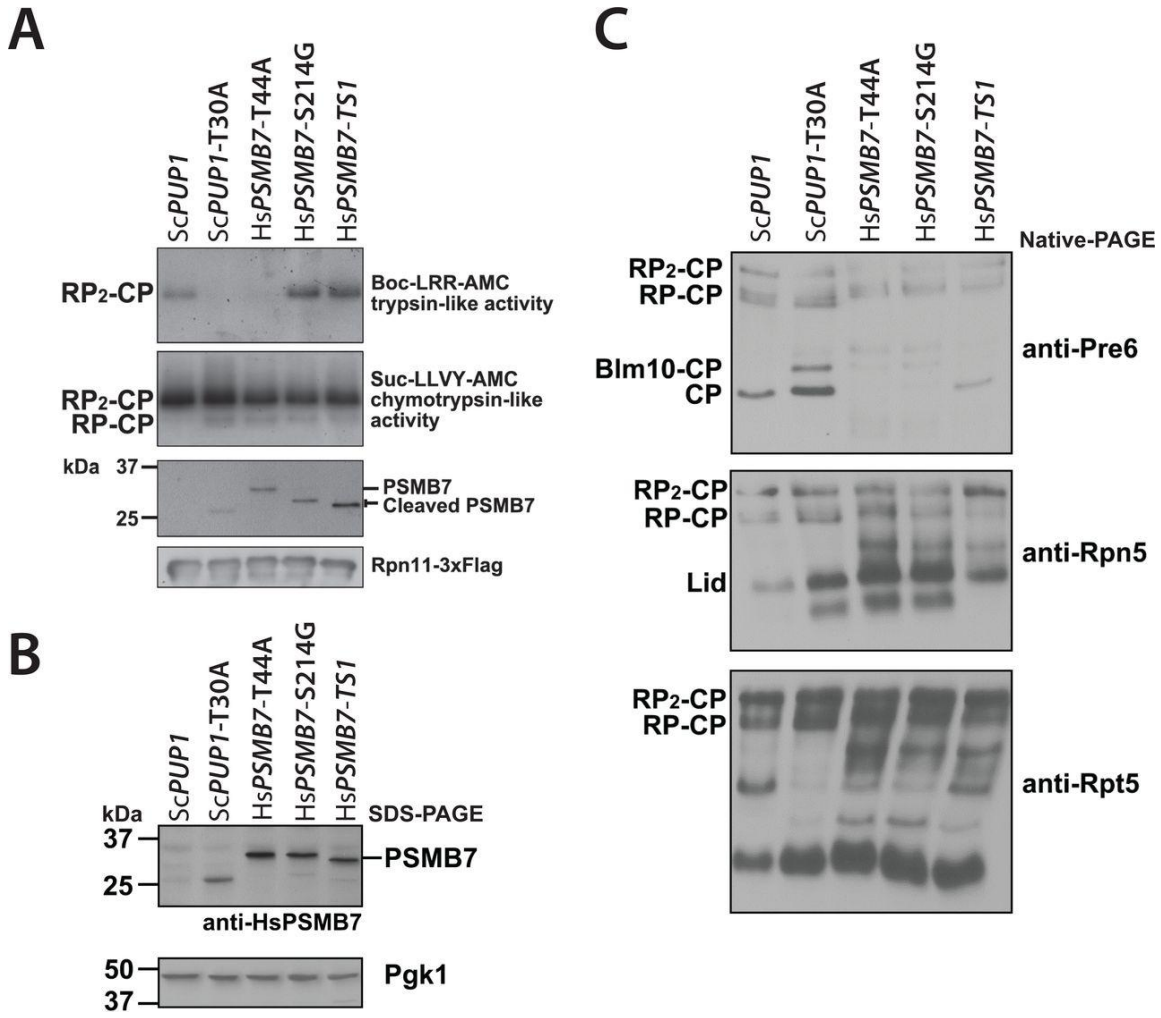


Figure 3.6 Complementing human *β*2 (PSMB7) variants in yeast reveal defects in proteasome assembly or PSMB7 auto processing.

(A) Yeast *pup1Δ* strains bearing plasmids expressing wild-type *ScPUP1*, *ScPUP1-T30A, HsPSMB7-T44A, HsPSMB7-S214G,* or *HsPSMB7-TS1* and harboring chromosomally tagged *RPN11-3xFLAG* were used to affinity purify 26S proteasomes via the 3xFLAG tag. The purified proteasomes were fractioned by native gel electrophoresis, and trypsin-like and chymotrypsin- like (from the *β*5 subunit) activities were tested using overlays with the fluorogenic Boc-LRR-AMC and Suc-LLVY-AMC substrates, respectively (top two panels). RP_2-CP and RP-CP are doubly and singly capped 26S proteasomes, respectively. The purified proteasomes were also separated by SDS-PAGE and analyzed by anti-*PSMB7* immunoblotting. Anti-PSMB7 antibody might weakly cross-react with Sc*Pup1*; the band in lane 2 might

therefore represent an unprocessed or partially processed *Pup1*, but this has not been verified. The anti-FLAG blot shows similar loading of samples. (B) Whole-cell lysates obtained from the same yeast strains as in panel (A) were directly fractionated by SDS-PAGE and analyzed by anti-PSMB7 immunoblotting. An anti-Pgk1 immunoblot was used as a loading control. (C) Native-PAGE immunoblot analyses of the cell lysates from panel (B) using antibodies to the indicated CP (Pre6/α4), lid (Rpn5), and base (Rpt5) subunits.

3.2.5 Wild-type human *β*2 can incorporate into yeast proteasomes if human *β*3 is also present.

Ideally, humanized yeast for functional characterization of human genes would utilize the wild-type human alleles in yeast. Our suppressor screen and the C-terminal tail swap data suggest that human *β2* requires compatible interactions with neighboring subunits, particularly *β3*, to assemble properly into the yeast CP. Thus, a strategy that restores these human-human subunit interactions might enable the integration of the wild-type PSMB7 into yeast proteasomes. To test this hypothesis, we asked if the wild type human *β2* can functionally replace its yeast counterpart in a strain that also expresses the human *β3* subunit.

We first tested a CRISPR-Cas9 methodology for deleting the yeast *β2* and *β3* genes. Plasmids expressing Cas9-sgRNA$^{Sc\beta2}$ or Cas9-sgRNA$^{Sc\beta3}$ were lethal, as expected (Figure S3.4A). Co-transformation of human gene repair templates harboring homology at the 5' and 3' termini of the corresponding yeast loci would be predicted to yield viable cells if the human genes could functionally replace the yeast orthologs. We first tested whether wild-type *HsPSMB7* could replace *ScPUP1* at the native locus but failed to obtain any viable colonies, as expected (**Figure 3.7A**). However, use of *HsPSMB7* variants, such as those with a T44A or S214G mutation, as repair templates allowed functional replacement of the yeast *PUP1* gene (**Figure 3.7B**). Previously, we had demonstrated that the yeast *β3* (*PUP3*) gene is functionally replaceable by its human ortholog (*HsPSMB3*) when expressed on a plasmid (Kachroo *et al.*, 2015). Using the CRISPR-Cas9-based strategy of HDR (Akhmetov *et al.*, 2018), we successfully replaced the genomic yeast *β3* gene with

human *β3* (*HsPSMB3*) (**Figure 3.7C**). Thus, the yeast *β*2 subunit can recruit human *β*3 to the yeast proteasome core, but human *β*2 is unable to do this with yeast *β*3.

Starting with the humanized *β3* yeast strain, CRISPR-Cas9-based genome editing now enabled replacement of yeast *β2* with the wild-type human *β2* (Hs*PSMB7*) gene. The doubly-humanized *Hsβ2-Hsβ3* yeast strain was viable, as verified by locus-specific PCR and Sanger sequencing (**Figure 3.7D**). Thus, by providing its neighboring human subunit, i.e., human *β*3, the normally non-complementing human *β*2 could now function in the yeast proteasome. Quantitative growth assays revealed modest fitness defects of the engineered strains. The humanized Hs*β*2-S214G strain showed a cold-sensitive growth defect at 23°C, whereas the humanized Hs*β*2-T44A strain grew slower at 23°C & 37°C compared to the wild-type strain. While the humanized *Hsβ3* strain grew comparably to wild-type yeast, the *Hsβ2-Hsβ3* strain manifested a cold-sensitive phenotype at 23°C and grew slower at 30°C in liquid culture (Figure S3.4B)

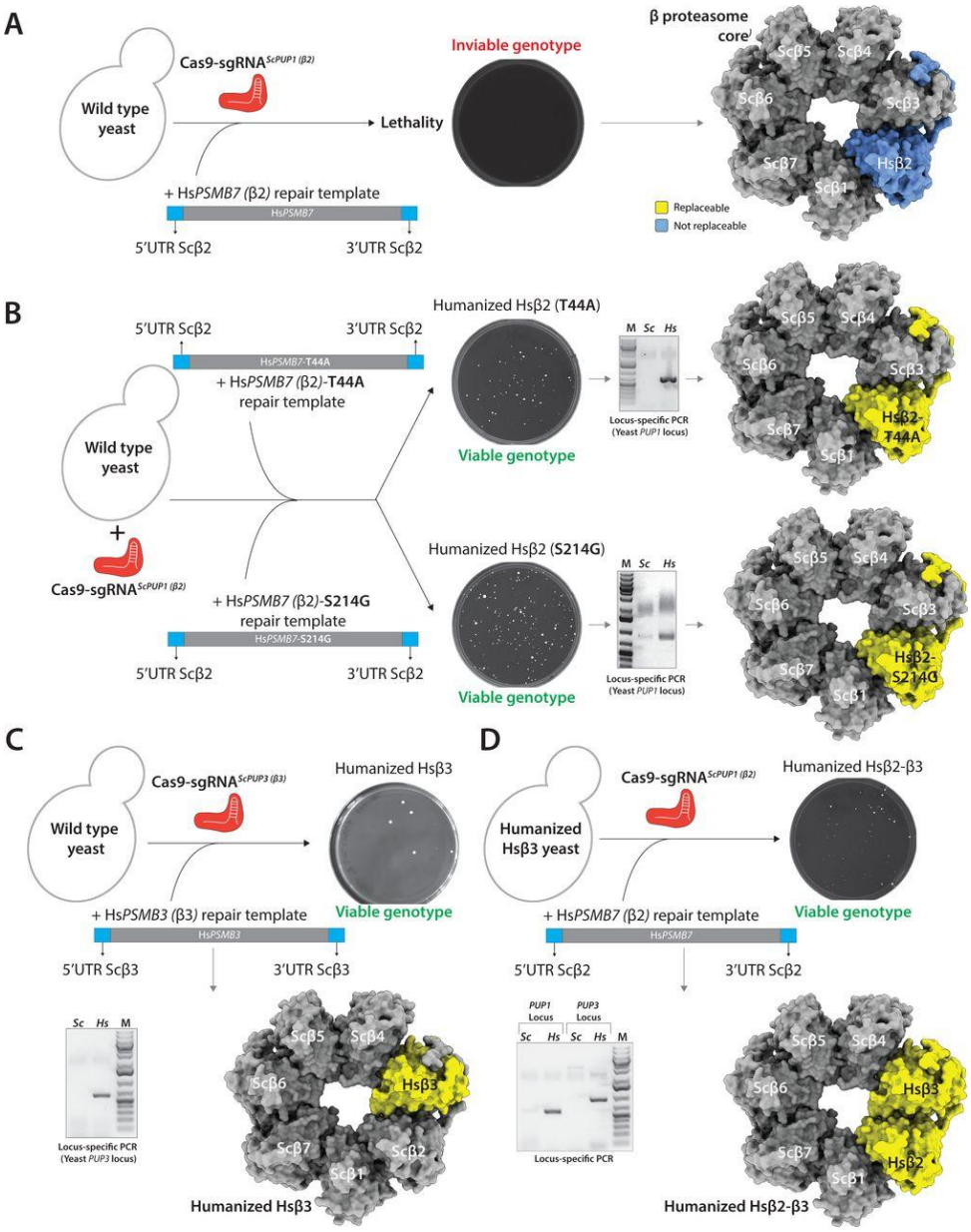


Figure 3.7 Restoration of neighboring interactions enables functional replaceability of wild-type human *β2* in yeast.

(A) Co-transformation of the pCas9-sgRNA^{ScPUP1} and human wild-type *PSMB7* repair template as a PCR fragment fails to obtain viable humanized strains. (B) However, the co-transformation of pCas9-sgRNA^{ScPUP1}and human *PSMB7-T44A* or *PSMB7-S214G* variants as a repair template yield viable yeast with genomically integrated human gene variants. (C) Transformation of the pCas9-sgRNA^{ScPUP3} and human wild-type PSMB3 (*β*3) repair template as a PCR fragment yields viable humanized *β*3 strains. (D) Using humanized *β*3 strains as a background, the co- transformation of pCas9-sgRNA^{ScPUP1} and human wild-type *PSMB7* as a repair template yield viable yeast with genomically integrated wild-type human *β2-β*3 genes in yeast. A single yeast *β*- proteasome core ring of 7-subunits is shown (PDB-1RYP) using ChimeraX software. The functionallly replaceable human subunits are indicated as yellow and the non-replaceable sunits are shown in blue.

3.3 Conclusion

Using an automated high-throughput pipeline and a large pool of mutant human *PSMB7*genes, we identified human *β2* variants that enable functional replacement of the yeast *β2 (Pup1)* ortholog. The variants reveal amino acids and protein domains that are critical for assembling human *β2* into the yeast proteasome core. All amino acid substitutions, except T44A, in human *PSMB7* that allow replacement of yeast *Pup1* are at or near interfaces with neighboring subunits, suggesting the contribution of multiple PPIs to the assembly of the core particle (CP). Several variants appear to promote the interaction with the yeast *β3* subunit. Modeling of the human *β2* in a fully assembled CP indicate that the mutations affect residues within an internal loop (A70V, D71N*) or C-terminal tail (K249R, E263K* and E236-Stop*) of *β*2 that are critical for interaction with the yeast *β*3 (**Figure 3**)**.** The C-terminal domain of yeast *β*2 is known to help guide the ordered assembly of the *β*-ring (Ramos *et al.*, 2004). However, the C-terminal tail- domains have diverged across species, suggesting the evolution of unique species-specific contacts. We show that replacing the C-terminal tail of human *β2* with the tail from yeast *β2* enables functional replaceability in yeast (**Figure3.5**). Alternatively, by providing a human-like PPI interface for *β2* by co-expression of the human *β3* subunit in yeast, the wild-type *β2/PSMB7* can functionally replace its yeast ortholog (**Figure3.7**). The data reveal that restoring local PPIs either via mutations or by providing a humanized neighbor enables an otherwise replacement-incompetent human gene to complement the orthologous yeast gene function (Huber *et al.*, 2016).

We assessed the functional replacement of yeast *β2* by human *PSMB7* variants via growth assays and biochemical characterization. While the human *β2* variants rescue the lethal growth defect of the knockout of the yeast ortholog, the humanized proteasome shows assembly defects based on the accumulation of distinct assembly intermediates and altered *β*2 precursor processing. These

defects are more pronounced in the T44A variant compared to the S214G and TS1 alleles. The trypsin-like catalytic activity of the Hs*PSMB7*-S214G and Hs*PSMB7*-TS1 variants is also comparable to that of wild-type yeast *PUP1* proteasomes, whereas the Hs*PSMB7*- T44A variant is catalytically inactive and accumulates in precursor form within proteasomes (**Figure3.6**)

Surprisingly, the catalytically dead *PSMB7*-T44A permits functional replaceability in yeast. This observation was unexpected. The structure of the 13S assembly intermediate shows an uncleaved propeptide of yeast *β*2 interacting with yeast *β*3 and apparently aiding in assembly (Schnell *et al.*, 2021). We show that the human *PSMB7*-T44A retains its propeptide, which might assist in the assembly of human *β*2 into yeast core particle. Since wild-type *β*2 will also have the propeptide until autocleavage at the very end of CP assembly after two half-mers have come together (Chen & Hochstrasser, 1996), the retained propeptide in the *PSMB7*-T44A mutant might stabilize *β*2-*β*3 within the preholo proteasome or help align the two half-mers. The human *β*2 propeptide is known to play a vital role in the cooperative assembly of the human *β*-ring, unlike the shorter propeptide of the yeast ortholog (De *et al.*, 2003; Tanaka, 2009; Budenholzer *et al.*, 2017; Hirano *et al.*, 2008). Together, our data suggest that the human *β*2 propeptide, despite its sequence divergence, has retained the ability to interact with the yeast *β*3 subunit (**Figure3.4**).

Our biochemical analysis has shown that several of the replacement-competent human *β*2 variants are proteolytically active. Further characterization of the incompatibilities associated with non-replaceable human *β* subunits should reveal a path to full humanization of the yeast proteasome core while identifying divergent ortholog functions. Yeast with a humanized catalytically active proteasome core provide a synthetic setup to characterize proteasome functionality *in vivo*. The strategy should enable the generation of distinct types of human proteasome cores (i.e., constitutive- and immuno-proteasomes) (Huber *et al.*, 2012; Nathan *et al.*, 2013), allowing the characterization

of their functions in a simplified cellular context. Proteasomes play a vital role in maintaining protein homeostasis and are implicated in human diseases ranging from cancer to age-related neurodegenerative disorders (Almond et al., 2002; Manasanch et al., 2017; Vilchez *et al.*, 2014). Although an attractive drug target, only a handful of FDA- approved drugs that inhibit the proteasome core are available (Chen *et al.*, 2011), (Park *et al.*, 2018), and no compounds are known that increase human CP catalytic activity (Njomen et al., 2019). Yeast with a humanized functionally active proteasome core provide a unique platform for discovering novel therapeutics to inhibit or enhance proteasome activity (Xin *et al.*, 2019). The humanized proteasome in yeast will enable direct assays deciphering human variant effects and gene-drug interactions that might alter proteasome function. Such tools should be useful for stratifying patients for different therapies - a step towards 'personalized medicine'.

3.4 Materials and methods

3.4.1 Constructing a *PSMB7* mutant gene library in a yeast expression vector

The *PSMB7* mutant gene library was previously generated (Kachroo *et al.*, 2015) by error-prone PCR (GeneMorph II Random Mutagenesis Kit from Agilent) to introduce mutations and add attL1 and attL2 sites at the 5' and 3' ends of the gene (Reece-Hoyes et al., 2018). The library was cloned using the LR cloning strategy into the expression vector pAG416GPD-ccdB (*URA3*; *CEN6*) where the *PSMB7* alleles are under the control of the constitutive GPD (*TDH3*) promoter (Gateway® LR Clonase® II enzyme mix kit from Invitrogen). The conditions for the error-prone PCR were selected to introduce 1-4 mutations per Kbp.

3.4.2 Transformation and selection of replaceable human gene suppressors (HsPSMB7) in yeast

Competent cells for the heterozygous knockout (HetKO) yeast (*PUP1*/*pup1*Δ*::kanMX*) with a Magic marker selection were made using the Frozen-EZ Yeast Transformation II Kit (Zymo Research). For maximum representation of mutant human gene clones transformed in the strain, the transformation was performed in larger scale (4X) than conventional methods and the transformation mix was plated on Q-trays (Corning, 245 mm Square BioAssay Dish) containing synthetic medium [SD-Ura with G418 (200 μg/ml)]. The trays were incubated at 30°C for 2-3 days. Next, >1000 single yeast colonies were picked using the QPix 460 colony picker. The single colonies were spotted on pre-sporulation GNA medium (5% glucose, 3% Difco nutrient broth, and 1% Difco yeast extract) with G418 selection (200 μg/ml) in a 96-spot format and incubated at 30°C for 1-2 days.

To select for viable haploid yeast knockout strains, each colony from pre-sporulation GNA medium was inoculated in 700 μL of liquid sporulation medium (0.1% potassium acetate (Sigma P1190), 0.005% zinc acetate (Sigma Z0625) in 96-well deep well plates. The mutant clones were incubated at room temperature (22-24°C) for 3-5 days while vigorously shaking at 230 rpm or by using a rotator. After confirming sporulation by brightfield microscopy, the spore mixes were plated on synthetic Magic Marker (MM) medium [-His -Arg -Leu -Ura +Can (60 μg/mL) with (yeast gene absent) or without G418 (yeast gene present) (200 μg/ml)] in a 96-well format and incubated at 30°C for 3-5 days. To further the growth of haploid yeast harboring replacement-competent Hs*PSMB7* mutants that grew similarly on MM-G418 and MM+G418, their corresponding spore mixes were diluted (1:20 dilution), plated on MM-G418 and MM+G418 petri plates, and incubated at 30°C for 3-5 days to obtain single colonies.

3.4.3 Human gene plasmid dependency assays

To test the human gene dependency of viable haploid knockout yeast (*pup1*Δ*::kanMX*), the haploid spores that grew on MM+G418 medium were replica plated on synthetic medium containing 5-FOA (1g/1L) from Thermo Fisher and uracil (50 mg/L) from Sigma Aldrich and incubated at 30°C for 1-2 days. Cells that did not grow on the 5-FOA medium were judged to be dependent on the plasmid-borne human gene variants for viability.

3.4.4 Plasmid preparation from yeast and *E. coli*

Plasmids harboring the mutant human *PSMB7* genes that passed the 5-FOA human gene dependency test were extracted from the original diploid HetKO strains. The cells were inoculated in YPD+G418 medium and incubated overnight at 30°C. The plasmids were extracted from yeast the following day using the QIAprep® Spin Miniprep kit. The plasmid yield from yeast was low. Therefore, the plasmids were transformed into *E. coli* and extracted from the resulting colonies using the QIAprep® Spin Miniprep kit.

3.4.5 Construction of Hs*PSMB7* single-site mutants via site-directed mutagenesis

PSMB7 single mutants were created with the use of the Q5® Site-Directed Mutagenesis kit from New England BioLabs. The primers for this experiment were designed using the software Geneious. We first created the wild-type human *PSMB7* entry clone in pDONR221 using the BP Gateway strategy followed by Sanger sequencing. Using the Q5 Site-Directed Mutagenesis kit, primers were used to introduce specific single-nucleotide changes in the wild-type *PSMB7*gene cloned in the pDONR221 entry clone. The forward primer introduced a mutation and in combination with a

compatible reverse primer the entire plasmid with a human gene was amplified. The linear plasmids were then treated with three enzymes from the kit: Kinase, Ligase, and DpnI to obtain circular human single mutant clones. The *PSMB7* single mutant clones were verified by restriction enzyme digestion (*EcoRV* and *HindIII* from New England BioLabs) followed by Sanger sequencing. The confirmed clones were moved into the yeast expression vector pAG416GPD-ccdB (*URA3; CEN6*) using the Gateway® LR Clonase® II enzyme mix kit.

3.4.6 Quantitative yeast growth assays

Yeast cells were inoculated in liquid SD-URA+G418 and grown overnight at 30°C. The yeast culture was inoculated in SD-URA+G418 at the initial OD (600 nm) of 0.01. The growth assay was performed with Biotek Synergy H1 plate reader for 48-72 hrs. while continuously shaking at 282 cpm and measuring the OD^{600} at 20-minute intervals. The OD^{600} measurements were then plotted to obtain growth curves for comparison using Graph Prism software.

3.4.7 Construction of human-yeast tail swap *HsPSMB7* clones in a yeast expression vector

Pymol-based structural evaluation, multiple-sequence alignments and human gene suppressor mutations were used to identify the C-terminal tails of the human and yeast β2 proteins. A common forward primer and three different reverse primers, each harboring a part of the *Pup1* C- terminal tail loop region (3' region) and a part of the *PSMB7* gene (5' region), were designed and used to create three different human-yeast hybrid genes by PCR (*AccuPrime Pfx* DNA polymerase from Invitrogen). The primers also add the attB1 and attB2 sites at the 5' and 3' ends of the PCR for cloning in pDONR221 entry clones using the BP cloning strategy. The unique reverse primers used for the construction of the tail-swap human *PSMB7* mutants are listed in **Table S1 (Supplementary**

file 1). The clones were confirmed by restriction digestion using *EcoRV* and *HindIII* and sequence verified. The verified hybrid human gene-yeast tail swap variants were cloned into the yeast expression vector pAG416GPD-ccdB (*URA3; CEN6*) by LR cloning (Gateway® LR Clonase® II enzyme mix kit from Invitrogen).

3.4.8 CRISPR-Cas9-based strategy to introduce a 3xFLAG tag at the endogenous *RPN11* gene

To design synthetic guide (sg) RNAs targeting the yeast proteasome *RPN11* gene, we used a built-in gRNA design tool in Geneious software (Kearse et al..,2012). We selected two guides with high ON-target and low OFF-target scores (**Figure S5**). The sgRNAs were synthesized as complementary oligos (IDT). After annealing the oligos, the 5' and 3' overhangs match the type IIS enzyme sites in a yeast expression vector pCAS9-*GFP*do with sgRNA expression system (CEN6, G418) (Lee *et al.*, 2015; Akhmetov *et al.*, 2018). See **Table S1 (Supplementary file 1)** for guide sequences and primers. Each sgRNA was cloned in the pCAS9-*GFP*do expression vector using the golden gate strategy. The plasmid allows the expression of an sgRNA, Cas9 nuclease, and an auxotrophic (URA3) or antibiotic selection (Geneticin-Sigma) marker.

The Golden Gate reaction for cloning sgRNAs was performed in a 10 µl volume, with approximately 20 fmol of annealed primer, 1 ul each of the *BsaI* FD (fast digestion, Thermo) enzyme, 1 µl of *T7 DNA ligase* (NEB), and 1 µl of *ATP* (NEB), 1ul of FD buffer (Thermo) and water to make up the volume. The Golden Gate reaction was performed in a PCR machine according to a previously published protocol (Akhmetov *et al.*, 2018). The reaction mix was transformed into competent *E. coli* cells and plated on LB agar with kanamycin (50 µg/ml). Since the sgRNA primers replace the *GFP* expression cassette, the correct clones were selected by screening for non-fluorescent colonies and verified by Sanger sequencing. The repair template for

RPN11-3xFLAG was synthesized as a gblock (IDT) with golden gate enzyme sites and cloned in pYTK001 and sequence-verified. The repair template was designed to harbor silent DNA sequence changes that allow efficient cloning in a vector using a Golden Gate reaction strategy (eliminates an internal enzyme site) while also carrying mutations in the sgRNA binding sites such that the engineered strains become resistant to DSBs (Double-Strand Break) by CRISPR/Cas9 .

Clones were initially screened using colony PCR and Phire plant direct master mix (Thermo). The forward primer for PCR screening was designed such that it binds outside of the ORF and homology used for HDR insertion of the repair template, and the reverse primer was designed to bind within the 3xFLAG tag. Following plasmid loss, clones were further verified by Sanger sequencing the entire *RPN11* locus using primers outside of the homology used for HDR.

3.4.9 CRISPR-Cas9-based genome editing to introduce wild-type human genes or their variants at the corresponding native yeast loci

The sgRNAs to target yeast *PUP1* and *PUP3* loci were generated using Geneious and cloned in pCAS9-*GFP*do as described above. See **Table S1 (Supplementary file 1)** for guide sequences and primers. The human gene repair templates for wild-type *HsPSMB7* (Hs*β*2 subunit) and *HsPSMB3* (Hs*β*3) were synthesized as a gblock with unique type IIs enzyme sites capable of generating distinct 4-base overhangs (IDT). To add native yeast locus homologous sequences, we amplified the 5'UTR (~500 bp) and 3'UTR (~150 bp) sequences of yeast *PUP1* and *PUP3* loci. The primers used to amplify the UTRs also harbor type IIs enzyme sites to clone (in YTK001) the UTRs with the corresponding human gene repair template gblocks using Golden Gate reactions as described above. The clones were sequence-verified, BsaI digested and directly used as repair templates for HDR. For Hs*PSMB7* variants, primers were designed with 80bp homology to the 5'

and 3' UTRs of the yeast loci. The high fidelity Accuprime enzyme was used to amplify variants from pAG416GPD expression constructs. The PCR product was gel extracted and used as a repair template for HDR in yeast.

3.4.10 Western blotting to test expression of *RPN11*-3xFLAG tag in yeast

Yeast strains were grown to mid-exponential phase in YPD medium to ~0.6 OD^{600}. The mixture was centrifuged for 3 minutes at 500 rpm followed by washing with 20 µl of 100 µM Tris-HCl (pH 8.0) containing a protease inhibitor cocktail (Millipore Sigma). The whole cell lysate was prepared by adding 200 µl of microbeads to the pellet and vortexed for 45 seconds followed by incubating on ice for 30 seconds. The procedure was repeated 5-8 times. The mixture was centrifuged at max speed (~10,000 rpm, eppendorf mini spin centrifuge) for 5 minutes. Approximately 20 µl of Laemmli sample buffer was added to the 100 µl of centrifuged supernatant and incubated at 65°C for 20 minutes. Twenty µl of each sample were loaded on precast SDS-PAGE gels (Thermo) and transferred to activated 0.2 µm PVDF membrane (Millipore Sigma) for Western blotting. RPN11-3xFLAG tagged subunits were detected by monoclonal anti-FLAG antibodies (Monoclonal ANTI-FLAG® M2 antibody #F3165, Millipore Sigma). The lysate from a wild-type strain with untagged *RPN11* was used as a negative control. For secondary detection, an anti-mouse antibody was used (IRDye 800CW Goat anti-Mouse IgG Secondary Antibody, LI-COR) and imaged by Odyssey (LI-COR Odyssey- 9120).

3.4.11 Yeast transformation

Yeast transformations were performed using the Frozen EZ Yeast Transformation II Kit from Zymo Research transformation according to the manufacturer's protocol.

3.4.12 Non-denaturing PAGE and SDS-PAGE of whole cell extracts

Yeast whole cell extracts were prepared as described previously (Li & Hochstrasser, 2022). Mid-exponential phase yeast cells were washed twice with ice-cold sterile water and frozen in liquid nitrogen. The frozen cells were ground with mortar and pestle, and the resulting cell powder was thawed on ice and resuspended in an equal volume of extraction buffer (50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 10% glycerol, 5 mM ATP). Extracts were centrifuged for 10 min at 21,000 x *g* to remove cell debris. Protein concentrations were determined using a Pierce BCA protein assay kit (Thermo Scientific, catalog # 23225, lot # SJ256254) according to the manufacturer's protocol. Equal amounts of total protein per sample (50 µg, except 10 µg for Pgk1) were loaded onto 4% native polyacrylamide gel electrophoresis (PAGE) gels and resolved for 3 hrs. at 100 V at 4°C for proteasome profile analysis. Alternatively, samples were resolved by 10% SDS-PAGE for analysis of Hs*PSMB7* expression and processing. Native PAGE-separated proteins or SDS PAGE-separated proteins were then transferred to PVDF membranes (EMD Millipore, catalog #IPVH00010, lot #R1EB02212) and subjected to Western blotting analysis as described previously (Li *et al.*, 2016) with the following primary antibodies: rabbit anti-Pre6 (Jäger *et al.*, 2001) at 1:5000 dilution, rabbit anti-Rpn5 (a generous gift from Daniel Finley lab at Harvard University) at 1:5000 dilution, rabbit anti-Rpt5 (Enzo Life Sciences, catalog # PW8245, lot # Z01946) at 1:10,000 dilution, rabbit anti-Hs*PSMB7* (Novus Biologicals, catalog # NBP2-19954, lot # 40275) at 1:1000 dilution, or an anti-Pgk1 monoclonal antibody (Abcam, catalog #ab113687, lot #GR3373682-5) at 1:10,000 dilution. Primary antibody binding was followed by anti-mouse-IgG (GE Healthcare, catalog #NXA931V, lot #17193521) or anti-rabbit-IgG (GE Healthcare, catalog #NA934V, lot #17212129) secondary antibody conjugated to horseradish peroxidase at the same dilution used with the primary

antibodies. The membranes were incubated in ECL detection reagent (Mruk & , 2011), and the ECL signals were detected using film (Thomas Scientific, catalog #1141J52).

3.4.13 Affinity purification of proteasomes and proteasome activity assay

Proteasomes were affinity purified from yeast cells expressing RPN11-3xFLAG as described previously (Li & Hochstrasser, 2022). Briefly, about 7 ml of cell powder from the same samples as used above for non-denaturing PAGE analysis were thawed on ice, resuspended in 10 ml buffer A (50 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol, 5 mM $MgCl_2$, 5 mM ATP, Roche complete EDTA-free protease inhibitor: catalog # 11873580001, lot # 53418500), and incubated for 15 min on ice. Cell debris was pelleted at 30,000 x *g* for 20 min at 4°C. Total protein concentrations of the supernatants were determined using the BCA assay. Supernatant equivalent to ~100 mg total protein was incubated with 200 µl (packed) resin of anti-FLAG M2 affinity gel (Sigma, catalog # A2220, lot # SLCH0130) for 2 h on a rotator at 4°C. The proteasome- bound resin was washed twice with 12 ml buffer A for 10 min, and then incubated with 3 resin volumes of 200 $\mu g \cdot ml^{-1}$ 3xFLAG peptide (Sigma, catalog # F4799, lot # SLCJ4916) for 45 min to elute proteasome complexes. Proteasomes were concentrated with 100K MWCO centrifugal filters (Merk Millipore, catalog # UFC510024, lot # R1MB60377) and quantified with a BSA standard using a G:Box Chemi HR16 imager (Syngene).

Proteasome activity analyses were performed as previously described (Li *et al.*, 2015), with minor modifications. Ten microgram purified proteasomes were loaded onto 4% native PAGE gels and resolved for 3 hrs. at 100 V at 4°C. The gels were incubated with developing buffer (50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 10% glycerol, 1 mM ATP) containing 50 µM substrates Boc-LRR-AMC (Enzo Life Sciences, catalog # BML-BW8515-0005, lot # 07232103) for trypsin-like activity

or Suc- LLVY-AMC (Sigma, catalog # S6510, lot # BCBK9233V) for chymotrypsin-like activity analysis with a gentle shaking at 30 rpm for 30 min at 30°C. Gels were transferred to a UV trans-illuminator and exposed to 365 nm light in a G:Box Chemi HR16 imager (Syngene).

3.4.13 Structural analysis

Homology modeling of the human *PSMB7* (Hsβ2; from PDB-1IRU) in the yeast proteasome core (PDB-1RYP) was performed uising Pymol [The Pymol molecular graphics system, version 2.5.2, Schrodinger, LLC.]. The structures to show yeast proteasome core and humanized Hsβ2 & Hsβ3 were designed using ChimeraX (Pettersen *et al.*, 2021).

3.5 Supplementary Figures

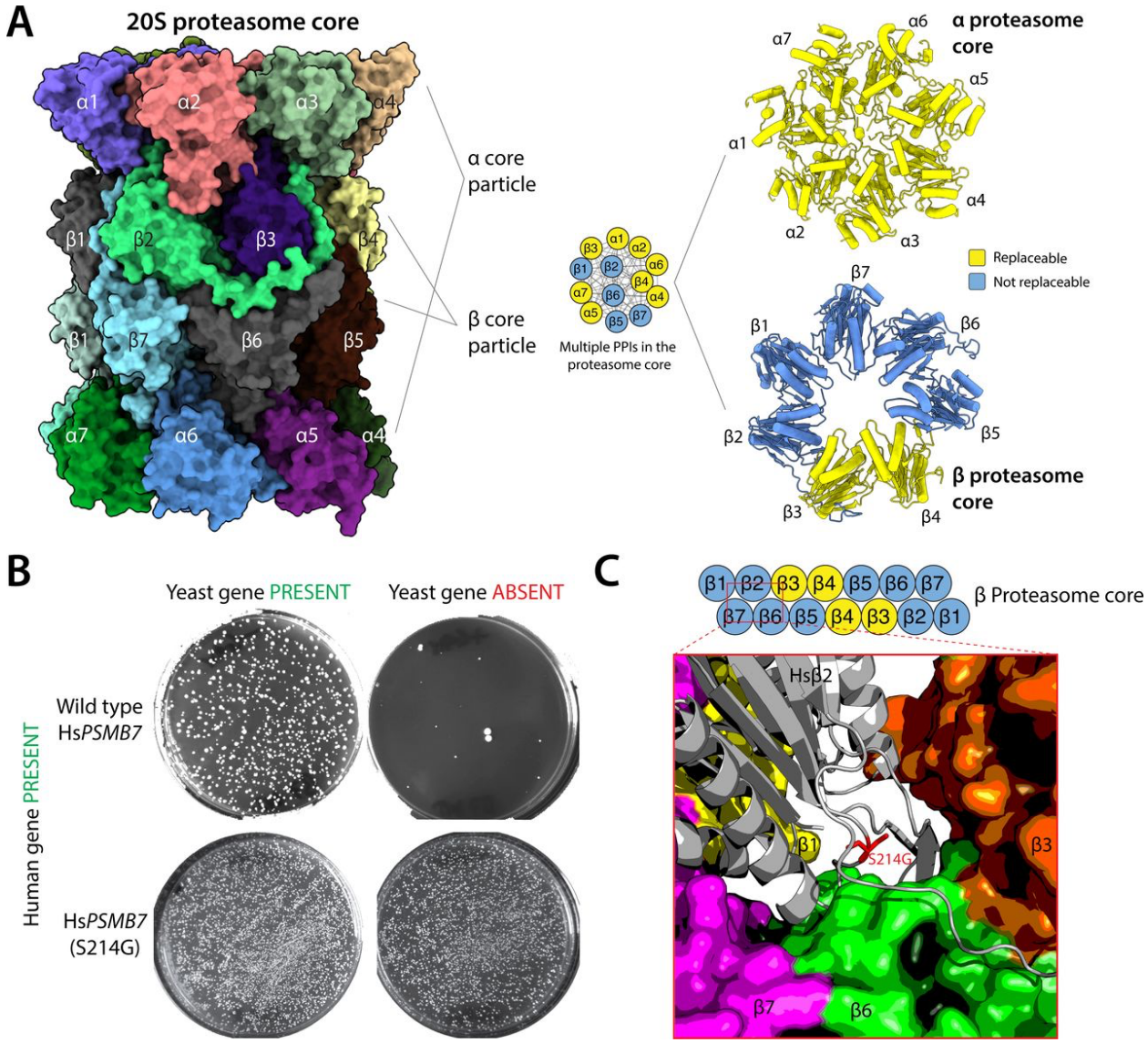


Figure S3.1. Functional replacement of yeast *β* subunits by human versions.

(A) A 28-subunit yeast proteasome core consists of two 7-subunit *a*-rings covering two 7-subunit *β*-rings. The structure shows many protein-protein interactions (PPIs) among the subunits to assemble the complex. Yeast proteasome core subunits are differentially replaceable by human orthologs (Blue indicates non-replaceable & Yellow indicates replaceable subunits). All *a* subunits are singly humanizable in yeast. However, 5 of 7 yeast *β* subunits cannot be replaced by their human equivalents. The images were generated using ChimeraX from PDB-1RYP. (B) A single Ser214Gly (S214G) mutation in the human *β*2 proteasome subunit (Hs*PSMB7*) is sufficient to allow it to assemble into

the yeast proteasome. Yeast strains lacking the yeast β2 (*PUP1*) gene are not complemented by wild-type human *PSMB7* (Yeast gene absent condition while few surving colonies represent segregation defects). However, the *PSMB7*-S214G mutant complements the deletion of yeast β2 (Yeast gene absent, human gene present condition). The rescue is similar to that seen with yeast β2 (Yeast gene present condition). (C) Structural modeling shows that the mutant human β2 residue (S214) is close to the interaction surface of the yeast β6 subunit, suggesting that restoration of the interaction is critical for replaceability. Human β2 (PDB-1IRU) was modeled in the yeast core proteasome structure [Images were generated using Pymol and human *PSMB7* or Hs β2 from PDB-1IRU modeled in the yeast proteasome core PDB-1RYP].

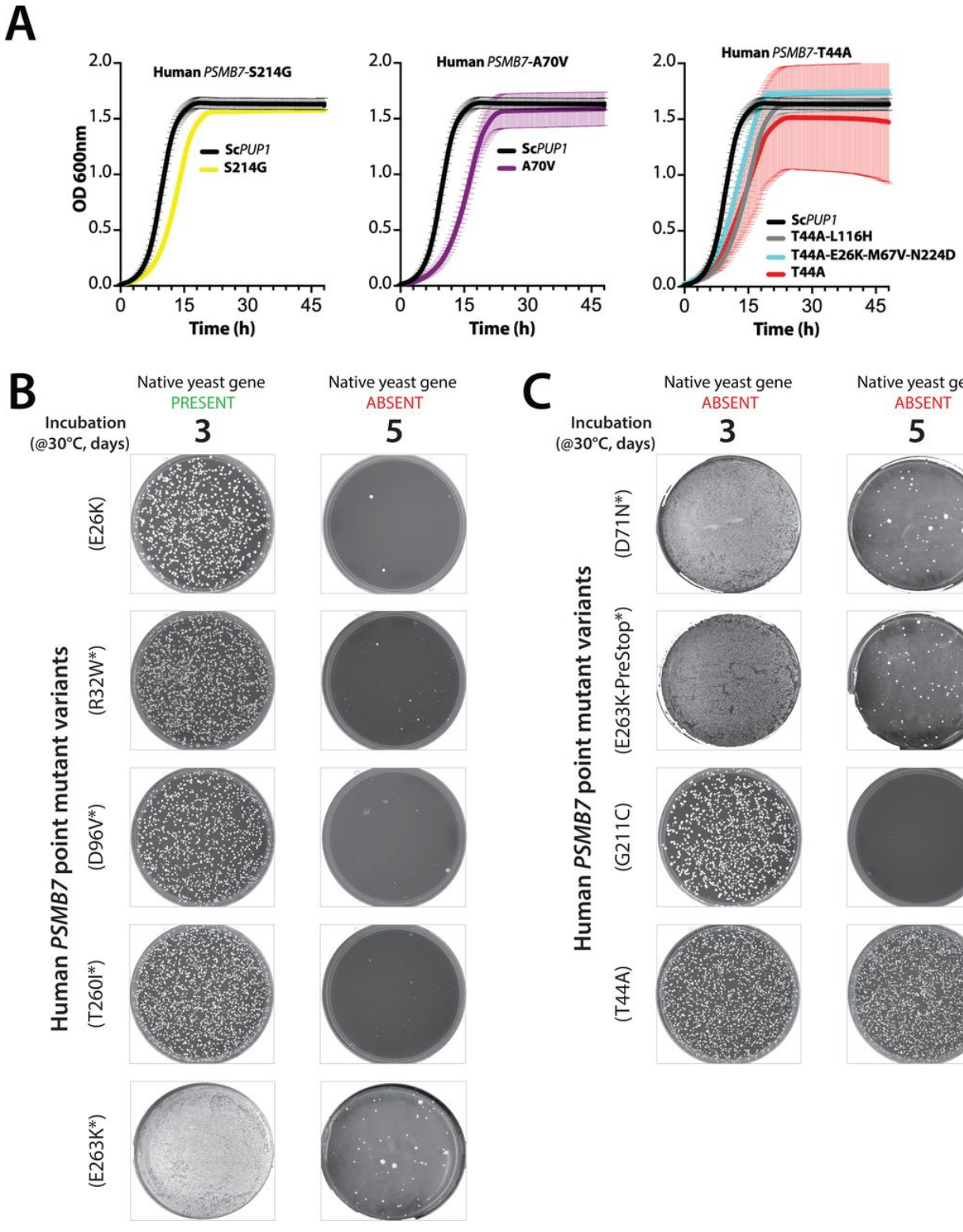

A
Human PSMB7-S214G
ScPUP1
S214G
OD 600nm
Time (h)
Human PSMB7-A70V
ScPUP1
A70V
Time (h)
Human PSMB7-T44A
ScPUP1
T44A-L116H
T44A-E26K-M67V-N224D
T44A
Time (h)
B
Native yeast gene
PRESENT
Native yeast gene
ABSENT
Incubation
(@30°C, days)
3
5
Human PSMB7 point mutant variants
(E26K)
(R32W*)
(D96V*)
(T260I*)
(E263K*)
C
Native yeast gene
ABSENT
Native yeast gene
ABSENT
Incubation
(@30°C, days)
3
5
Human PSMB7 point mutant variants
(D71N*)
(E263K-PreStop*)
(G211C)
(T44A)

Figure S 3.2. Quantitative growth assays to compare the replaceability of single-site mutants to their primary suppressors.

(A) Quantitative growth profiles of humanized yeast strains harboring Hs*PSMB7*-S214G (yellow), Hs*PSMB7*-A70V (purple), or Hs*PSMB7*-T44A (red) variants [compared to the primary suppressors Hs*PSMB7*-T44A-L116H (gray), Hs*PSMB7*-T44A-E26K-M67V-N224D (blue)] or the positive control yeast *PUP1* (black). (B) Post-sporulation selection to grow haploids on Magic Marker medium (MM) with G418 (Yeast gene ABSENT) or without G418 (Yeast gene PRESENT) enables selection for functional replaceability. The expression of yeast *PUP1* under the control of the constitutive GPD promoter functionally complements the growth defect of *pup1Δ::kanMX* strain (within 3-6 days of incubation at 30°C), whereas the empty vector shows no growth. The assays tested several single-site mutants in *HsPSMB7* associated with primary suppressors, showing no functional replaceability phenotype in yeast. (C) Previously obtained single-site mutants in *HsPSMB7* (Kachroo et al. , 2015) derived from primary suppressors bearing multiple mutations cannot replace yeast *PUP1*, in contrast to Hs*PSMB7*-T44A, which shows robust colony formation after 3-4 days of incubation at 30°C.

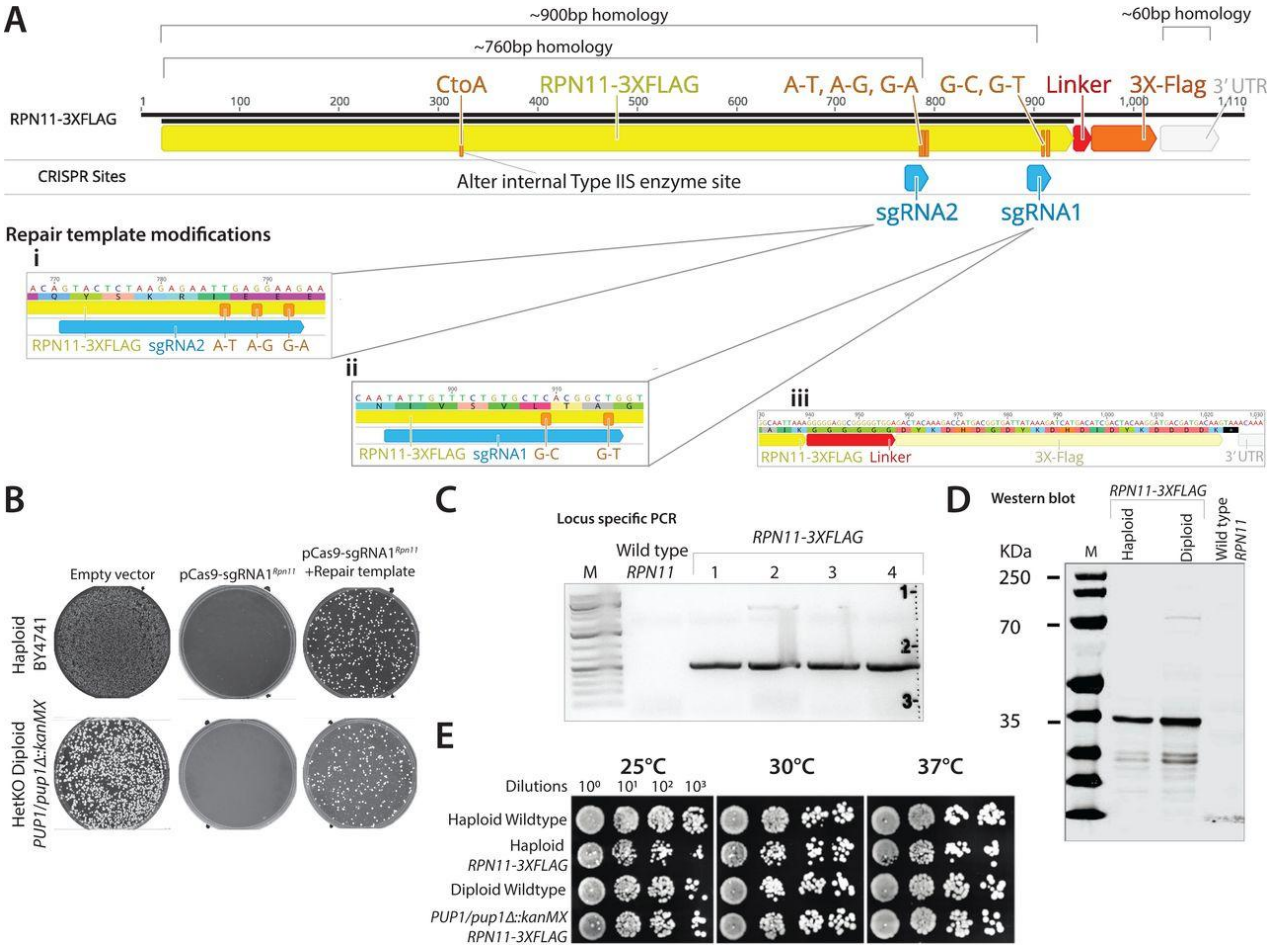


Figure S3.3 Genomic insertion and characterization of a 3xFLAG tag at the C-terminal coding sequence of the yeast *RPN11* RP gene.

(A) sgRNAs targeting the yeast *RPN11* locus are shown as cyan arrows. The sgRNA target sites (i and ii) had high ON-target and low OFF-target scores with sgRNA target (ii) closer to the edit site (i.e. 3xFlag) than site (i). The repair template (*RPN11*-3xFLAG) carries silent mutations at the sgRNA sites. This ensures that after CRISPR-Cas9 induced double-strand break and repair via homologous recombination, the edited site becomes resistant to further targeting by the Cas9- sgRNA complex. (B)The Cas9-sgRNA1^{RPN11} expression is lethal in wild-type haploid or heterozygous knockout diploid (*PUP1/pup1Δ::kanMX)* yeast carrying a wild-type *RPN11* locus. The co-transformation of the CRISPR reagent and the repair template permits the survival of many yeast colonies. The successful edits at the *RPN11* locus were verified by (C) locus-specific PCR, and (D) Western blotting (using anti-FLAG antibody) to confirm the expression of *RPN11*- 3xFLAG protein. (E) Quantitative growth assays performed as dilutions on solid agar media at various temperatures (25°C, 30°C and 37°C) show no apparent fitness defects due to the 3xFLAG tag sequence at the *RPN11* locus in either haploid or diploid (*PUP1/pup1Δ::kanMX*; with both copies of *RPN11* carrying a 3xFlag) yeast strains.

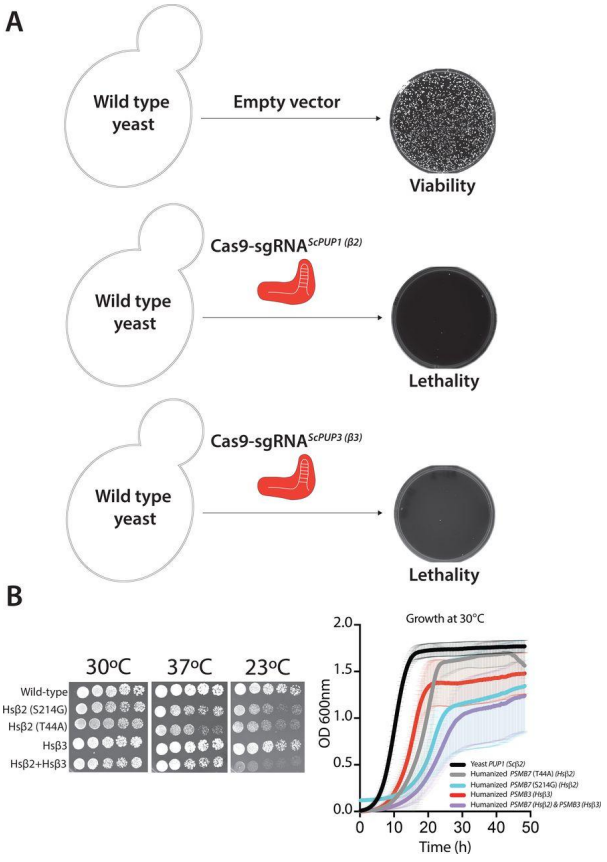


Figure S3.4 CRISPR-Cas9 based genomic replacement of human *PSMB7* (Hs*β*2) and human PSMB3 (Hs*β*3) in yeast.

(A) The expression of Cas9-sgRNA reagents targeting yeast *PUP1* (pCas9-sgRNA^{ScPUP1}) or *PUP3* (pCas9-sgRNA^{ScPUP3}) loci show lethality compared to the Cas9 alone. (F) Growth assays performed using yeast liquid culture (YPD) at 30°C and serial dilutions spotted onto solid agar media grown at various temperatures (30°C, 37°C and 23°C) of genomically replaced wild-type human *β*2, its variants and human *β*3 in yeast. Shown are growth rates of humanized yeast with genomically replaced yeast genes [Hs*PSMB7*-S214G (cyan), Hs*PSMB7*-T44A (gray), Hs*PSMB3* or Hs*β3* (red), and Hs*PSMB3* + Hs*PSMB7* or Hs*β2-β3* strains] compared to the wild type yeast (black).

3.6 Supplementary Table S1

Table S3.1Primers used in this study.

Name	Primer sequence (5' to 3')	use
PSMB7-attB1-Fp	GATCACAAGTTTGTACAAAAAAGCAGGCTTCATGGCG GCTGTGTCGGTGT	Add attB1 site to the 5' end of Hs*PSMB7*
PSMB7-attb2-TS1-Rp	GATCACCACTTTGTACAAGAAAGCTGGGTCAAGCCGTT ATATCGACTTGTTCTTCTTGTATGTCACAAATATTGACA ATACTCTCCTTCAGCACAGCAGTTGTACCCCTGGGGAA TTTGTAGCTTTTCTGCTTTTCTTCTCTAACATTTGGAGTC AAGTATGGGCGGAGAAAATCCAGC	Add attB2 site and Full-length C-terminal tail of Sc*PUP1* to Hs*PSMB7* at the 3' end
PSMB7-attB2-TS2-Rp	GATCACCACTTTGTACAAGAAAGCTGGGTCAAGCCGTT ATATCGACTTGTTCTTCTTGTATGTCACAAATATTGACA ATACTCTCCTTCAGCACAGCAGTTGTACCCCTCTCACAC CTGTACCGGCCAA	Add attB2 site and Full-length partial C-terminal tail of Sc*PUP1* to Hs*PSMB7* at the 3' end
PSMB7-attB2-TS3-Rp	GATCACCACTTTGTACAAGAAAGCTGGGTCAAGCCGTT ATATCGACTTGTTCTTCTTGTATGTCACAAATATTGACG ATTTTCTCAGTGAGGACTGCAGT	Add attB2 site and Full-length smallest C-terminal tail of Sc*PUP1* to Hs*PSMB7* at the 3' end
Sc*RPN11*-3xFlag-Fp	AGCCTTTAAGCTAAGAAACGGTT	For genotyping
Sc*RPN11*-3xFlag-R	GGTCTTTGTAGTCTCCACCCC	For genotyping
Sc*RPN11*-seq-650-Fp	TCAAGTTTACGATGGCGCCT	For sequencing
Sc*RPN11*-seq-140-Rp	AGCCTTTTCCCTCCTAACGC	For sequencing

Sc*RPN11*-sgRNA1	GTACTCTAAGAGAATAGAAG	gRNA target
Sc*RPN11*-sgRNA2	TATTGTTTCTGTGCTGACGG	gRNA target
Sc*PUP3*-sgRNA1	GATCCAAGTTCTATTAACGG	gRNA target
Sc*PUP3*-sgRNA2	ATTGCCTGTGATTTGCGTCT	gRNA target

Table S3.2. Plasmids used in this study.

Name	**Reference**
pYTKOO1	Lee et al. , 2015
pYTKOO1-*RPN11*-3xFLAG	This study
pDKO-Cas9-sgRNA-URA	This study
pDkO-Cas9-sgRNA-*kanMX*	This study
pCas9-sgRNA*RPN11*	This study
pCas9-sgRNA*PUP1*	This study
pCas9-sgRNA*PUP3*	This study
p416GPD-Sc*PUP1*	This study
p416GPD-Sc*PSMB7*	Kachroo et al. , 2015
p416GPD-Sc*PSMB7*-T44A	Kachroo et al. , 2015
p416GPD-Sc*PSMB7*-S214G	Kachroo et al. , 2015
p416GPD-Sc*PSMB7*-variants	This study
pYTK-*PSMB7*-ScUTR	This study
pYTK-*PSMB3*-ScUTR	This study

Table S3.3. Yeast strains used in this study.

Strain name	Genotype	Reference
BY4741	*MATa* his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Saccharomyces Genome Database
BY4741-*RPN11*-3xFLAG	BY4741; *RPN11*-3′tag-3xFLAG	This study
BY4743-Magic Marker strain-*RPN11*-	BY4743; *MATa*/*αPUP1*/*pup1*Δ::*kanMX*; *RPN11*-3′tag-3xFLAG, CAN1/can1Δ::LEU2-MFA1pr-HIS3	This study
BY4741-Hs*β*2 (T44A)	BY4741; *PUP1* Δ ::Hs*PSMB7*(T44A)	This study
BY4741-Hs*β*2 (S214G)	BY4741; *PUP1* Δ::Hs*PSMB7*(S214G)	This study
BY4741-Hs*β*3	BY4741; *PUP3* Δ::HsPSMB3	This study
BY4741-Hs*β*2*β*3	BY4741; *PUP1* Δ ::Hs*PSMB7*; *PUP3*Δ::HsPSMB3	This study

Chapter 4

Conclusions and future directions

In this thesis, I describe a CRISPR-Cas9 based genome editing strategy to humanize nearly the entire yeast alpha proteasome core. The regulation by the native yeast promoter and terminator conveniently provides a more accurate test of functional replaceability. The efficiency and speed of our CRISPR/Cas9 method allows multiplexed and rapid editing in yeast without the need to undergo laborious selection/counter-selection cycles.

Owing to the haploid and diploid selection of yeast, and using CRISPR-Cas9 gene drives, we developed a method combining multiple genetic edits in a single yeast strain. This Marker less, Enrichment and Recombination of Genetically Engineered Loci (MERGE) involves a mating distinct genetically engineered strains followed by transformation with CRISPR plasmids and to screen for strains harboring multi-gene edited (humanized) loci. We explore this approach to generate a near complete human alpha proteasome core in yeast.

An important hypothesis regarding the functional replacement of human genes in yeast is the concept of genetic modularity *i.e.,* functional replaceable (or non-replaceable) human genes tend to have local physical or genetic interactions with genes that are functional replaceable as well. The modularity paradigm, therefore, suggests that individually non-replaceable proteins may be become replaceable if the neighboring human interaction partners are also provided. To test this hypothesis, the β-subunit ring of the proteasome core provides an ideal platform. Since, the β-subunits interact with each other, the neighboring humanizable α-ring, and the assembly chaperones, the following strategies were explored regarding the successful replaceability of the human β subunits:

1. If the expression of the entire human β-subunits module is functionally replaceable in yeast
2. If the assembly of human β-subunits require a fully humanized α-ring

The co-expression of human constitutive- and immuno-βs followed by MS analysis in the singly- and partially humanized strains showed stable expression of human β subunits only in partially humanized yeast strains suggesting likely incorporation in the proteasome core in partly humanized α-ring.

As an alternate strategy, we developed a novel pipeline to screen for human *PSMB7* gene (Hs β2) suppressors that allow functional replaceability of the orthologous *PUP1* gene in yeast. The strategy discovered point mutations in human *PSMB7* that enabled functional replaceability. Several point mutations were found in the C-terminal tail of Hs β2. Exchanging the full-length Hsβ2 with that of Sc*PUP1* enabled complementation. Finally, we showed that co-expression of the Hsβ3 subunit in yeast, enables functional complementation of the wild type β2/*PSMB7*. Thus, by restoring local physical interactions (or PPIs) either via mutations or by providing a humanized neighbor enables functional replaceability. In future, a similar pipeline can be used to humanize the rest of non-replaceable human β core genes.

Humanizing β ring

Humanizing the entire β ring is challenging, as we don't know which interfaces are incompatible. We have already generated strains with β3 and β4 genomically replaced, one of the strategies could replace the rest of Hsβs in the partially humanized Hsβ3/ β4 background. Given that the β subunits follow a regulated but well-defined pathway of assembly onto the scaffold that the α-ring (Murata et al., 2009), this well-defined assembly pathway should aid us in deciding the order in which to build the humanized β ring. Alternatively, β subunits can be inserted in heterozygous diploids or at the landing pad loci followed by MERGE strategy to functionally replace all human βs in yeast..

Humanizing proteasome assembly chaperones

The assembly of the proteasome requires a set of extrinsic proteasome assembly chaperones and intrinsic chaperones like N-terminal pro-peptides and C-terminal tails of β subunits. Human and yeast catalytic β subunits (β1, β2 and β5) along with β6 and β7 are synthesized with unprocessed N-terminal pro-peptides which upon correct assembly of CP are cleaved. In addition, several accessory factors are required for optimal and correct sequential assembly of CP. While highly conserved, species-species specific changes may have altered assembly intermediates in human and yeast proteasomes. Together, these studies reveal the differences between mammalian and yeast proteasome core assemblies suggesting, at least, at the level of chaperone recognition, these protein-protein interactions are orthogonal in mammals and yeast. There are several interacting chaperone type proteins that regulate core particle assembly (Hanssum et al.., 2014; Murata et al.., 2009). These can easily be added from human, either exogenously or by humanizing their yeast orthologs, to help facilitate core assembly, particularly of the β-ring.

Humanized proteasome in yeast explores new proteasome biology.

Humanizing the yeast proteasome provides a novel and unique platform to study the diverged properties of the human proteasome in a simple cellular context. The regulation of proteasome homeostasis in the context of disease is an emerging field and the humanized yeast will open new avenues to discover novel, yet uncharacterized, modes of proteasome regulation, assembly and nuclear transport, providing yeast reagents to manipulate/study individual human proteasome subunits for therapeutic purposes. Proteasome mutations are clinically linked to several human diseases (Cavo, M et al., 2006). Yeast with humanized proteasomes will help interpret the

contribution of human genetic variation on proteasome function and drug action and illuminate the principles of genetic disease, leading to quantitative models for errors of human proteostasis. Strains harboring partly or fully humanized pathways will be used for assaying combination therapies and will serve as biological test tubes for identifying important human variants by deep scanning mutagenesis (Fowler, Stephany, & Fields, 2014). Bortezomib is the first proteasome inhibitor used in the treatment of multiple myeloma, thus exploring more novel inhibitors in the humanized strains could provide promising anticancer therapies. Studying interaction of drugs with variants of humanized subunits either singly or in combination could be a valuable therapeutic strategy to impede proteasome function for personalized medicine.

Bibliography

1. Agmon, N., Temple, J., Tang, Z., Schraink, T., Baron, M., Chen, J., Mita, P., Martin, J. A., Tu, B. P., Yanai, I. (2020). Phylogenetic debugging of a complete human biosynthetic pathway transplanted into yeast. Nucleic Acids Res. 48, 486-499.

2. Ahn, J., Choi, C.-H., Kang, C.-M., Kim, C.-H., Park, H.-M., Song, K.-B., Hoe, K.-L., Won, M. and Chung, K.-S. (2009). Generation of expression vectors for high- throughput functional analysis of target genes in *Schizosaccharomyces pombe*. J. Microbiol. 47, 789-795.

3. Aiyar, R. S., Bohnert, M., Duvezin-Caubet, S., Voisset, C., Gagneur, J., Fritsch, E. S., Couplan, E., von der Malsburg, K., Funaya, C., Soubigou, F. (2014). Mitochondrial protein sorting as a therapeutic target for ATP synthase disorders. Nat. Commun. 5:5585.

4. Akhmetov, A., Laurent, J. M., Gollihar, J., Gardner, E. C., Garge, R. K., Ellington, A. D., Kachroo, A. H. and Marcotte, E. M. (2018). Single-step precision genome editing in yeast using CRISPR-Cas9. Bio. Protoc. 8, e2765.

5. Alberts, B. (2010). Model organisms and human health. Science 330, 1724.

6. Alliance of Genome Resources Consortium. (2019). The alliance of genome resources: building a modern data ecosystem for model organism databases. Genetics 213, 1189-1196.

7. Alliance of Genome Resources Consortium. (2020). Alliance of genome resources portal: unified model organism research platform. Nucleic Acids Res. 48, D650-D658.

8. Almond JB & Cohen GM (2002) The proteasome: a novel target for cancer chemotherapy. Leukemia 16: 433–443.

9. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 43, D789-D798.

10. Amorosi, C. J., Chiasson, M. A., McDonald, M. G., Wong, L. H., Sitko, K. A., Boyle, G., Kowalski, J. P., Rettie, A. E., Fowler, D. M. and Dunham, M. J. (2021).

11. Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. Am. J. Hum. Genet. 108, 1735-1751.

12. Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285, 901-906.

13. Arendt CS & Hochstrasser M (1997) Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. Proc Natl Acad Sci U S A 94: 7156–7161.

14. Aylon, Y. and Kupiec, M. (2004). New insights into the mechanism of homologous recombination in yeast. Mutat. Res. 566, 231-248.

15. Bach, S., Talarek, N., Andrieu, T., Vierfond, J.-M., Mettey, Y., Galons, H., Dormont, D., Meijer, L., Cullin, C. and Blondel, M. (2003). Isolation of drugs active against mammalian prions using a yeast-based screening assay. Nat. Biotechnol. 21, 1075-1081.

16. Balakrishnan, R., Park, J., Karra, K., Hitz, B. C., Binkley, G., Hong, E. L., Sullivan, J., Micklem, G. and Michael Cherry, J. (2012). YeastMine—an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose toolkit. Database 2012, bar062.

17. Barbieri, E. M., Muir, P., Akhuetie-Oni, B. O., Yellman, C. M. & Isaacs, F. J. (2017).Precise Editing at DNA Replication Forks Enables Multiplex Genome Engineering in Eukaryotes. Cell 171, 1453–1467.e13.

18. Balasov, M., Akhmetova, K. and Chesnokov, I. (2020). Humanized Drosophila model of the Meier-Gorlin syndrome reveals conserved and divergent features of the Orc6 protein. Genetics 216, 995-1007.

19. Bellen, H. J., Wangler, M. F. and Yamamoto, S. (2019). The fruit fly at the interface of diagnosis and pathogenic mechanisms of rare and common human diseases. Hum. Mol. Genet. 28, R207-R214.

20. Bellen, H. J., Hubbard, E. J. A., Lehmann, R., Madhani, H. D., Solnica-Krezel, L. and Southard-Smith, E. M. (2021). Model organism databases are in jeopardy. Development 148, dev200193.

21. Bolognesi, B., Faure, A. J., Seuma, M., Schmiedel, J. M., Tartaglia, G. G. and Lehner, B. (2019). The mutational landscape of a prion-like domain. Nat. Commun. 10, 4162.

22. Boone, C. (2014). Yeast systems biology: our best shot at modeling a cell. Genetics 198, 435-437.

23. Botstein, D. and Fink, G. R. (1988). Yeast: an experimental organism for modern biology. Science 240, 1439-1443.

24. Botstein, D. and Fink, G. R. (2011). Yeast: an experimental organism for 21st century biology. Genetics 189, 695-704.

25. Bourgeois, L., Pyne, M. E. & Martin, V. J. J. (2018). A Highly Characterized Synthetic Landing Pad System for Precise Multicopy Gene Integration in Yeast. ACS Synth Biol 7, 2675–2685.

26. Budenholzer L, Cheng CL, Li Y & Hochstrasser M (2017) Proteasome Structure and Assembly. J Mol Biol 429: 3500–3524.

27. Brzá č ová, Z., Peť ková , M., Veljač iková , K., Zajič ková , T. and Tomá š ka, Ľ . (2021). Reconstruction of human genome evolution in yeast: an educational primer for use with "systematic humanization of the yeast cytoskeleton discerns functionally replaceable from divergent human genes.". Genetics 219, iyab118.

28. Chan, K.-M., Liu, Y.-T., Ma, C.-H., Jayaram, M. and Sau, S. (2013). The 2-micron plasmid of Saccharomyces cerevisiae: a miniaturized selfish genome with optimized functional competence. Plasmid 70, 2-17.

29. Cherry, J. M. (2015). The saccharomyces genome database: a tool for discovery. Cold Spring Harb. Protoc. 2015.

30. Chen D, Frezza M, Schmitt S, Kanwar J & Dou QP (2011) Bortezomib as the first proteasome inhibitor anticancer drug: current status and future perspectives. Curr Cancer Drug Targets 11: 239–253.

31. Chen P & Hochstrasser M (1996) Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. Cell 86: 961–972.

32. Costanzo, M., Kuzmin, E., van Leeuwen, J., Mair, B., Moffat, J., Boone, C. and Andrews, B. (2019). Global genetic networks and the genotype-to-phenotype relationship. Cell 177, 85-100.

33. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823

34. Couplan, E., Aiyar, R. S., Kucharczyk, R., Kabala, A., Ezkurdia, N., Gagneur, J., St. Onge, R. P., Salin, B., Soubigou, F., Le Cann, M. (2011). A yeast-based assay identifies drugs active against human mitochondrial disorders. Proc. Natl Acad. Sci. USA 108, 11989-11994.

35. De M, Jayarapu K, Elenich L, Monaco JJ, Colbert RA & Griffin TA (2003) Beta 2 subunit propeptides influence cooperative proteasome assembly. J Biol Chem 278: 6153–6159.

36. DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J. and Church, G. M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res. 41, 4336-4343.

37. DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M. & Church, G. M. Safeguarding CRISPR-Cas9 gene drives in yeast (2015).. Nat Biotechnol 33, 1250–1255.

38. Dixon, S. J., Fedyshyn, Y., Koh, J. L. Y., Prasad, T. S. K., Chahwan, C., Chua, G., Toufighi, K., Baryshnikova, A., Hayles, J., Hoe, K.-L. (2008). Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. Proc. Natl. Acad. Sci. USA 105, 16653-16658.

39. Dolinski, K. and Botstein, D. (2007). Orthology and functional conservation in eukaryotes. Annu. Rev. Genet. 41, 465-507.

40. Dryja, T. P., Cavenee, W., White, R., Rapaport, J. M., Petersen, R., Albert, D. M. and Bruns, G. A. (1984). Homozygosity of chromosome 13 in retinoblastoma. N. Engl. J. Med. 310, 550-553. doi:10.1056/NEJM198403013100902.

41. Duina, A. A., Miller, M. E. and Keeney, J. B. (2014). Budding yeast for budding geneticists: a primer on the Saccharomyces cerevisiae model system. Genetics 197, 33-48. doi:10.1534/genetics.114.163188.

42. Dwight, S. S., Harris, M. A., Dolinski, K., Ball, C. A., Binkley, G., Christie, K. R., Fisk, D. G., Issel-Tarver, L., Schroeder, M., Sherlock, G. (2002). Saccharomyces genome database (SGD) provides secondary gene annotation using the gene ontology (GO). Nucleic Acids Res. 30, 69-72.

43. Fowler, D. M., Araya, C. L., Fleishman, S. J., Kellogg, E. H., Stephany, J. J., Baker, D. and Fields, S. (2010). High-resolution mapping of protein sequence- function relationships. Nat. Methods 7, 741-746.

44. Fowler, D. M., Stephany, J. J. and Fields, S. (2014). Measuring the activity of protein variants on a large scale using deep mutational scanning. Nat. Protoc. 9, 2267-2284.

45. Friedmann, Y., Shriki, A., Bennett, E. R., Golos, S., Diskin, R., Marbach, I., Bengal, E. and Engelberg, D. (2006). JX401, A p38α inhibitor containing a 4- benzylpiperidine motif, identified via a novel screening system in yeast. Mol. Pharmacol. 70, 1395-1405.

46. Fredens, J., Wang, K., de la Torre, D., Funke, L.F.H., Robertson, W.E., Christova, Y., Chia, T., Schmied, W.H., Dunkelmann, D.L., Beránek, V. (2019). Total synthesis of Escherichia coli with a recoded genome. Nature 569, 514–518.

47. Gabaldó n, T. and Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. Nat. Rev. Genet. 14, 360-366.

48. Garge, R. K., Laurent, J. M., Kachroo, A. H. and Marcotte, E. M. (2020). Systematic humanization of the yeast cytoskeleton discerns functionally replaceable from divergent human genes. Genetics 215, 1153-1169.

49. Garge, R. K., Cha, H. J., Lee, C., Gollihar, J. D., Kachroo, A. H., Wallingford, J. B. and Marcotte, E. M. (2021). Discovery of new vascular disrupting agents based on evolutionarily conserved drug action, pesticide resistance mutations, and humanized yeast. Genetics 219, iyab101.

50. Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A. (2008). Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319, 1215–1220.

51. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature 418, 387-391.

52. Giorgini, F., Guidetti, P., Nguyen, Q., Bennett, S. C. and Muchowski, P. J. (2005) A genomic screen in yeast implicates kynurenine 3-monooxygenase as a therapeutic target for Huntington disease. Nat. Genet. 37, 526-531.

53. Goffeau, A., Aert, R., Agostini-Carbone, M. L., Ahmed, A., Aigle, M., Alberghina, L., Albermann, K., Albers, M., Aldea, M., Alexandraki, D. (1997). . The yeast genome directory Nature 387, 5.

54. Golden, A. (2017). From phenologs to silent suppressors: identifying potential therapeutic targets for human disease. Mol. Reprod. Dev. 84, 1118-1132.

55. Groll M, Heinemeyer W, Jäger S, Ullrich T, Bochtler M, Wolf DH & Huber R (1999) The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. Proc Natl Acad Sci USA 96: 10976–10983.

56. Hamza A, Driessen MRM, Tammpere E, O'Neil NJ & Hieter P (2020) Cross-Species Complementation of Nonessential Yeast Genes Establishes Platforms for Testing Inhibitors of Human Proteins. Genetics 214: 735–747.

57. Harshbarger W, Miller C, Diedrich C & Sacchettini J (2015) Crystal structure of the human 20S proteasome in complex with carfilzomib. Structure 23: 418–424

58. Hartwell, L. H. (2004). Yeast and cancer. Biosci. Rep. 24, 523-544.

59. Hedges, S. B. (2002). The origin and evolution of model organisms. Nat. Rev. Genet. 3, 838-849.

60. Heinicke, S., Livstone, M. S., Lu, C., Oughtred, R., Kang, F., Angiuoli, S. V., White, O., Botstein, D. and Dolinski, K. (2007). The Princeton Protein Orthology Database (P-POD): a comparative genomics analysis tool for biologists. PLoS One 2, e766.

61. Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., Hoon, S., Lee, W., Proctor, M., St Onge, R. P., Tyers, M., Koller, D. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 320, 362-365.

62. Hirano, Y., Hayashi, H., Iemura, S.-I., Hendil, K.B., Niwa, S.-I., Kishimoto, T., Kasahara, M., Natsume, T., Tanaka, K., and Murata, S. (2006). Cooperation of multiple chaperones required for the assembly of mammalian 20S proteasomes. Mol Cell 24, 977–984.

63. Hirano Y, Kaneko T, Okamoto K, Bai M, Yashiroda H, Furuyama K, Kato K, Tanaka K & Murata S (2008) Dissecting beta-ring assembly pathway of the mammalian 20S proteasome. EMBO J 27: 2204–2213.

64. Hoffman, C. S., Wood, V. and Fantes, P. A. (2015). An ancient yeast for young geneticists: a primer on the schizosaccharomyces pombe model system. Genetics 201, 403-423.

65. Hohmann, S. (2016). Nobel yeast research. FEMS Yeast Res. 16, fow094.

66. Huber EM, Basler M, Schwab R, Heinemeyer W, Kirk CJ, Groettrup M & Groll M (2012) Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity. Cell 148: 727–738.

67. Huber EM, Heinemeyer W, de Bruin G, Overkleeft HS & Groll M (2016) A humanized yeast proteasome identifies unique binding modes of inhibitors for the immunosubunit β5i. EMBO J 35: 2602–2613.

68. Hunter, P. (2008). The paradox of model organisms: the use of model organisms in research will continue despite their shortcomings. EMBO Rep. 9, 717-720.

69. Huh, W.-K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature 425, 686–691

70. Ishioka, C., Frebourg, T., Yan, Y. X., Vidal, M., Friend, S. H., Schmidt, S. and Iggo, R. (1993). Screening patients for heterozygous p53 mutations using a functional assay in yeast. Nat. Genet. 5, 124-129.

71. Jäger S, Strayle J, Heinemeyer W & Wolf DH (2001) Cic1, an adaptor protein specifically linking the 26S proteasome to its substrate, the SCF component Cdc4. EMBO J 20: 4423–4431

72. Jo, M., Chung, A. Y., Yachie, N., Seo, M., Jeon, H., Nam, Y., Seo, Y., Kim, E., Zhong, Q., Vidal, M. (2017). Yeast genetic interaction screen of human genes associated with amyotrophic lateral sclerosis: identification of MAP2K5 kinase as a potential drug target. Genome Res. 27, 1487-1500.

73. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K.Bates, R.,Žıdek,A.,Potapenko,A.etal.(2021).Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589.

74. Kachroo AH, Laurent JM, Akhmetov A, SzilagyiJones M, McWhite CD, Zhao A & Marcotte EM (2017) Systematic bacterialization of yeast genes identifies a near-universally swappable pathway. Elife 6. 10.7554/eLife.25093.

75. Kachroo AH, Laurent JM, Yellman CM, Meyer AG, Wilke CO & Marcotte EM (2015) Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science 348: 921–925.

76. Kachroo AH, Vandeloo M, Greco BM & Abdullah M (2022) Humanized yeast to model human biology, disease and evolution. Dis Model Mech 15: dmm049309.

77. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649

78. Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 28, 617-623.

79. Kusmierczyk AR, Kunjappu MJ, Funakoshi M & Hochstrasser M (2008) A multimeric assembly factor controls the formation of alternative 20S proteasomes. Nat Struct Mol Biol 15: 237–244

80. Kucharczyk, R., Salin, B. and di Rago, J.-P. (2009). Introducing the human Leigh syndrome mutation T9176G into Saccharomyces cerevisiae mitochondrial DNA leads to severe defects in the incorporation of Atp6p into the ATP synthase and in the mitochondrial morphology. Hum. Mol. Genet. 18, 2889-2898.

81. Kumar, A. and Snyder, M. (2001). Emerging technologies in yeast genomics. Nat. Rev. Genet. 2, 302-312.

82. Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Mattiazzi Usaj, M., van Leeuwen, J. (2018). Systematic analysis of complex genetic interactions. Science 360, eaao1729.

83. Kuzmin, E., Costanzo, M., Andrews, B., and Boone, C. (2016). Synthetic Genetic Array Analysis.. Cold Spring Harb Protoc, pdb. prot088807.

84. Kramer, K. M., BROck, J. A., Bloom, K., Moore, J. K. & Haber, J. E. Two different types of double-strand breaks in Saccharomyces cerevisiae are repaired by similar RAD52-independent, nonhomologous recombination events (1994). Mol Cell Biol 14, 1293–1301.

85. Lamesch, P., Li, N., Milstein, S., Fan, C., Hao, T., Szabo, G., Hu, Z., Venkatesan, K., Bethel, G., Martin, P. (2007). hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics 89, 307-315.

86. Labunskyy, V.M., Suzuki, Y., Hanly, T.J., Murao, A., Roth, F.P., and Gladyshev, V.N. (2014). The insertion Green Monster (iGM) method for expression of multiple exogenous genes in yeast. G3 (Bethesda) 4, 1183–1191.

87. Laurent, J. M., Young, J. H., Kachroo, A. H. and Marcotte, E. M. (2016). Efforts to make and apply humanized yeast. Brief Funct. Genomics 15, 155-163.

88. Laurent JM, Garge RK, Teufel AI, Wilke CO, Kachroo AH & Marcotte EM (2020) Humanization of yeast genes with multiple human orthologs reveals functional divergence between paralogs. PLoS Biol 18: e3000627.

89. Lee ME, DeLoache WC, Cervantes B & Dueber JE (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth Biol 4: 975–986.

90. Le Tallec, B., Barrault, M.-B., Courbeyrette, R., Guérois, R., Marsolier-Kergoat, M.-C., and Peyroche, A. 2007). 20S proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and in mammals. Mol Cell 27, 660–674.

91. Li J, Fuchs S, Zhang J, Wellford S, Schuldiner M & Wang X (2016) An unrecognized function for COPII components in recruiting the viral replication protein BMV 1a to the perinuclear ER. J Cell Sci 129: 3597–3608.

92. Li J & Hochstrasser M (2022) Selective microautophagy of proteasomes is initiated by ESCRT-0 and is promoted by proteasome ubiquitylation. J Cell Sci 135: jcs259393.

93. Li Y, Tomko RJ & Hochstrasser M (2015) Proteasomes: Isolation and Activity Assays. Curr Protoc Cell Biol 67: 3.43.1–3.43.20.

94. Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H.-R., Haimovich, A.D., Kuznetsov, G., Mercer, J.A., Wang, H.H., Carr, P.A., Mosberg, J.A. (2013). Genomically recoded organisms expand biological functions. Science 342, 357–360.

95. Lee, M. G. and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. Nature 327, 31-35.

96. Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.

97. Leonelli, S. and Ankeny, R. A. (2012). Re-thinking organisms: the impact of databases on model organism biology. Stud. Hist. Philos. Biol. Biomed. Sci. 43, 29-36.

98. Li, Z., Vizeacoumar, F. J., Bahr, S., Li, J., Warringer, J., Vizeacoumar, F. S., Min, R., Vandersluis, B., Bellay, J., Devit, M. (2011). Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat. Biotechnol. 29, 361-367.

99. Liao, B.-Y. and Zhang, J. (2008). Null mutations in human and mouse orthologs frequently result in different phenotypes. Proc. Natl. Acad. Sci. USA 105, 6987-6992.

100. õoke, M., Kristjuhan, K. from yeasts for PCR-based applications. Biotechniques 50, 325–328.

101. Lu, X. and Horvitz, H. R. (1998). lin-35 and lin-53, two genes that antagonize a C. elegans Ras pathway, encode proteins similar to Rb and its binding protein RbAp48. Cell 95, 981-991.

102. Manasanch EE & Orlowski RZ (2017) Proteasome inhibitors in cancer therapy. Nat Rev Clin Oncol 14: 417–433.

103. Marini, N. J., Gin, J., Ziegle, J., Keho, K. H., Ginzinger, D., Gilbert, D. A. and Rine, J. (2008). The prevalence of folate remedial MTHFR enzyme variants in humans. Proc. Natl. Acad. Sci. USA. 105, 8055-8060.

104. Maupin-Furlow J (2011) Proteasomes and protein conjugation across domains of life. Nat Rev Microbiol 10: 100–111.

105. Marini, N. J., Thomas, P. D. and Rine, J. (2010). The use of orthologous sequences to predict the impact of amino acid substitutions on protein function. PLoS Genet. 6, e1000968.

106. McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B. and Marcotte, E. M. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc. Natl. Acad. Sci. USA 107, 6544-6549.

107. McWhite, C. D., Liebeskind, B. J. and Marcotte, E. M. (2015). Applications of comparative evolution to human disease genetics. Curr. Opin. Genet. Dev. 35, 16-24.

108. MGC Project Team, Temple, G., Gerhard, D. S., Rasooly, R., Feingold, E. A., Good, P. J., Robinson, C., Mandich, A., Derge, J. G., Lewis, J., Shoaf, D. .. (2009). The completion of the Mammalian Gene Collection (MGC). Genome Res. 19, 2324-2333.

109. Mitchell, L.A., Chuang, J., Agmon, N., Khunsriraksakul, C., Phillips, N.A., Cai, Y., Truong, D.M., Veerakumar, A., Wang, Y., Mayorga, M. (2015). Versatile genetic assembly system (VEGAS) to assemble pathways for expression in S. cerevisiae. Nucleic Acids Res 43, 6620–6630.

110. Mnaimneh, S., Davierwala, A. P., Haynes, J., Moffat, J., Peng, W.-T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A. (2004). Exploration of essential gene functions via titratable promoter alleles. Cell 118, 31-44.

111. Morris, E. P. & da Fonseca, P. C. A. How to build a proteasome. (2021). Nat Struct Mol Biol 28, 409–410 .

112. Murata, S., Yashiroda, H. & Tanaka, K. (2009). Molecular mechanisms of proteasome assembly. Nat Rev Mol Cell Biol 10, 104–115.

113. Mruk DD & Cheng CY (2011) Enhanced chemiluminescence (ECL) for routine immunoblotting: An inexpensive alternative to commercially available kits. Spermatogenesis 1: 121–122.

114. Nathan JA, Spinnenhirn V, Schmidtke G, Basler M, Groettrup M & Goldberg AL (2013) Immuno- and constitutive proteasomes do not differ in their abilities to degrade ubiquitinated proteins. Cell 152: 1184–1194.

115. Nurse,P. (2002). Cyclin dependent kinases and cell cycle control (nobel lecture). Chembiochem 3, 596-603.

116. Njomen E & Tepe JJ (2019) Proteasome Activation as a New Therapeutic Approach To Target Proteotoxic Disorders. J Med Chem 62: 6469–6481.

117. Outeiro, T. F. and Lindquist, S. (2003). Yeast cells provide insight into alpha- synuclein biology and pathobiology. Science 302, 1772-1775.

118. Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, Hieter P, Spencer F & Boeke JD (2004) A robust toolkit for functional profiling of the yeast genome. Mol Cell 16: 487–496.

119. Park JE, Miller Z, Jun Y, Lee W & Kim KB (2018) Next-generation proteasome inhibitors for cancer therapy. Transl Res 198: 1–16.

120. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH & Ferrin TE (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci 30: 70–82.

121. Perkins, E., Sun, D., Nguyen, A., Tulac, S., Francesco, M., Tavana, H., Nguyen, H., Tugendreich, S., Barthmaier, P., Couto, J. (2001). Novel inhibitors of poly (ADP-ribose) polymerase/PARP1 and PARP2 identified using a cell-based screen in yeast. Cancer Res. 61, 4175-4183.

122. Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. 9, 855-867.

123. Prelich, G. (2012). Gene overexpression: uses, mechanisms, and interpretation. Genetics 190, 841-854.

124. Ramos PC & Dohmen RJ (2008) PACemakers of proteasome core particle assembly. Structure 16: 1296–1304.

125. Ramos PC, Marques AJ, London MK & Dohmen RJ (2004) Role of C-terminal extensions of subunits beta2 and beta7 in assembly and activity of eukaryotic proteasomes. J Biol Chem 279: 14323–14330.

126. Rajan, V., Melong, N., Hing Wong, W., King, B., Tong, S. R., Mahajan, N., Gaston, D., Lund, T., Rittenberg, D., Dellaire, G. (2020). Humanized zebrafish enhance human hematopoietic stem cell survival and promote acute myeloid leukemia clonal diversity. Haematologica 105, 2391-2399.

127. Rak, M., Tetaud, E., Godard, F., Sagot, I., Salin, B., Duvezin-Caubet, S., Slonimski, P. P., Rytka, J. and di Rago, J.-P. (2007).

128. Yeast cells lacking the mitochondrial gene encoding the ATP synthase subunit 6 exhibit a selective loss of complex IV and unusual mitochondrial morphology. J. Biol. Chem. 282, 10853-10864.

129. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8, 2281–2308.

130. Reece-Hoyes JS & Walhout AJM (2018) Gateway Recombinational Cloning. Cold Spring Harb Protoc 2018.

131. Remm M, Storm CE & Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314: 1041–1052.

132. Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., DiCarlo, J. E., Lee, D., Huang, C. L. V., Chandrasegaran, S., Cai, Y. (2017). Design of a synthetic yeast genome. Science 355, 1040-1044.

133. Roggenkamp, E., Giersch, R.M., Wedeman, E., Eaton, M., Turnquist, E., Schrock, M.N., Alkotami, L., Jirakittisonthon, T., Schluter-Pascua, S.E., Bayne, G.H (2017). CRISPR-UnLOCK: Multipurpose Cas9-Based Strategies for Conversion of Yeast Libraries and Strains. Front Microbiol 8, 1773.

134. Rine, J. (2014). A future of the model organism model. Mol. Biol. Cell 25, 549-553.

135. Roscito, J. G., Sameith, K., Parra, G., Langer, B. E., Petzold, A., Moebius, C., Bickle, M., Rodrigues, M. T. and Hiller, M. (2018). Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. Nat. Commun. 9, 4737.

136. Rut W & Drag M (2016) Human 20S proteasome activity towards fluorogenic peptides of various chain lengths. Biol Chem 397: 921–926.

137. Saito, T. L., Ohtani, M., Sawai, H., Sano, F., Saka, A., Watanabe, D., Yukawa, M., Ohya, Y. and Morishita, S. (2004). SCMD: saccharomyces cerevisiae morphological database. Nucleic Acids Res. 32, D319-D322.

138. Sakata E, Stengel F, Fukunaga K, Zhou M, Saeki Y, Förster F, Baumeister W, Tanaka K & Robinson CV (2011) The catalytic activity of Ubp6 enhances maturation of the proteasomal regulatory particle. Mol Cell 42: 637–649.

139. Scherens, B. and Goffeau, A. (2004). The uses of genome-wide yeast mutant collections. Genome Biol. 5, 229.

140. Sekigawa, M., Kunoh, T., Wada, S.-I., Mukai, Y., Ohshima, K., Ohta, S., Goshima, N., Sasaki, R. and Mizukami, T. (2010). Comprehensive screening of human genes with inhibitory effects on yeast growth and validation of a yeast cell- based system for screening chemicals. J. Biomol. Screen. 15, 368-378.

141. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498-2504.

142. Shashikant, C. S., Carr, J. L., Bhargava, J., Bentley, K. L. and Ruddle, F. H. (1998). Recombinogenic targeting: a new approach to genomic analysis—a review. Gene 223, 9-20.

143. Schnell, H.M., Walsh, R.M., Rawson, S., Kaur, M., Bhanu, M.K., Tian, G., Prado, M.A., Guerra-Moreno, A., Paulo, J.A., Gygi, S.P. (2021). Structures of chaperone-associated assembly intermediates reveal coordinated mechanisms of proteasome biogenesis. Nat Struct Mol Biol 28, 418–425.

144. Sirr, A., Lo, R. S., Cromie, G. A., Scott, A. C., Ashmead, J., Heyesus, M. and Dudley, A. M. (2020). A yeast-based complementation assay elucidates the functional impact of 200 missense variants in human PSAT1. J. Inher. Metab. Dis. 43, 758-769.

145. Smardová, J. (1999). FASAY: a simple functional assay in yeast for identification of p53 mutation in tumors. Neoplasma 46, 80-88.

146. Sonnhammer, E. L. L. and Ö stlund, G. (2015). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. 43, D234-D239.

147. Starita, L. M. and Fields, S. (2015). Deep mutational scanning: a highly parallel method to measure the effects of mutation on protein function. Cold Spring Harb. Protoc. 2015, 711-714.

148. Sun, S., Yang, F., Tan, G., Costanzo, M., Oughtred, R., Hirschman, J., Theesfeld, C. L., Bansal, P., Sahni, N., Yi, S. (2016). An extended set of yeast-based functional assays accurately identifies human disease mutations. Genome Res. 26, 670-680.

149. Stadt Mueller, B.M., Kish-Trier, E., Ferrell, K., Petersen, C.N., Robinson, H., Myszka, D.G., Eckert, D.M., Formosa, T., and Hill, C.P. (2012). Structure of a proteasome Pba1-Pba2 complex: implications for proteasome assembly, activation, and biological function. J Biol Chem 287, 37371–37382.

150. Sun, S., Weile, J., Verby, M., Wu, Y., Wang, Y., Cote, A. G., Fotiadou, I., Kitaygorodsky, J., Vidal, M., Rine, J. (2020). A proactive genotype-to- patient-phenotype map for cystathionine beta-synthase. Genome Med. 12, 13.

151. Suzuki, Y., St Onge, R.P., Mani, R., King, O.D., Heilbut, A., Labunskyy, V.M., Chen, W., Pham, L., Zhang, L.V., Tong, A.H.Y. (2011). Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. Nat Methods 8, 159–164.

152. Suzuki, Y., Stam, J., Novotny, M., Yachie, N., Lasken, R.S., and Roth, F.P. (2012) J Vis Expe4072.

153. Teufel, A. I., Johnson, M. M., Laurent, J. M., Kachroo, A. H., Marcotte, E. M. and Wilke, C. O. (2019). The many nuanced evolutionary consequences of duplicated genes. Mol. Biol. Evol. 36, 304-314.

154. Tanaka K (2009) The proteasome: overview of structure and functions. Proc Jpn Acad Ser B Phys Biol Sci 85: 12–36.

155. Tomko, R. J. and Hochstrasser, M. (2013). Molecular architecture and assembly of the eukaryotic proteasome. Annu. Rev. Biochem. 82, 415-445.

156. Truong, D. M. and Boeke, J. D. (2017). Resetting the yeast epigenome with human nucleosomes. Cell 171, 1508-1519.e13.

157. Tschumper, G. and Carbon, J. (1983). Copy number control by a yeast centromere. Gene 23, 221-232.

158. Tugendreich, S., Bassett, D. E., McKusick, V. A., Boguski, M. S. and Hieter, P. (1994). Genes conserved in yeast and humans. Hum. Mol. Genet. 3, 1509-1517.

159. UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49, D480-D489.

160. Van Leeuwen, J., Boone, C. and Andrews, B. J. (2017). Mapping a diversity of genetic interactions in yeast. Curr Opin. Syst. Biol. 6, 14-21.

161. VanderSluis, B., Costanzo, M., Billmann, M., Ward, H. N., Myers, C. L., Andrews, B. J. and Boone, C. (2018). Integrating genetic and protein-protein interaction networks maps a functional wiring diagram of a cell. Curr. Opin. Microbiol. 45, 170-179.

162. Vo, T. V., Das, J., Meyer, M. J., Cordero, N. A., Akturk, N., Wei, X., Fair, B. J., Degatano, A. G., Fragoza, R., Liu, L. G. (2016). A proteome-wide fission yeast interactome reveals network evolution principles from yeasts to human. Cell 164, 310-323.

163. Vilchez D, Saez I & Dillin A (2014) The role of protein clearance mechanisms in organismal ageing and age-related diseases. Nat Commun 5: 5659.

164. Waseem, N. H., Labib, K., Nurse, P. and Lane, D. P. (1992). Isolation and analysis of the fission yeast gene encoding polymerase delta accessory protein PCNA. EMBO J. 11, 5111-5120.

165. Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science285, 901–906.

166. Weile, J., Sun, S., Cote, A. G., Knapp, J., Verby, M., Mellor, J. C., Wu, Y., Pons, C., Wong, C., van Lieshout, N. (2017). A framework for exhaustively mapping functional missense variants. Mol. Syst. Biol. 13, 957.

167. Woods, J. O., Singh-Blom, U. M., Laurent, J. M., McGary, K. L. and Marcotte, E. M. (2013). Prediction of gene-phenotype associations in humans, mice, and plants using phenologs. BMC Bioinformatics 14, 203.

168. Xin BT, Huber EM, deBruin G, Heinemeyer W, Maurits E, Espinal C, Du Y, Janssens M, Weyburne ES, Kisselev AF, et al.. (2019) Structure-Based Design of Inhibitors Selective for Human Proteasome β2c or β2i Subunits. J Med Chem 62: 1626–1642.

169. Xu, H., Han, M., Zhou, S., Li, B.-Z., Wu, Y., and Yuan, Y.-J. (2020). Chromosome drives via CRISPR-Cas9 in yeast. Nat Commun 11, 4344.

170. Yang F, Sun S, Tan G, Costanzo M, Hill DE, Vidal M, Andrews BJ, Boone C & Roth FP (2017) Identifying pathogenicity of human variants via paralog-based yeast complementation. PLoS Genet 13: e1006779.

171. Yan, Y. & Finnigan, G. C. Development of a multi-locus CRISPR gene drive system in budding yeast (2018). Sci Rep 8, 17277).

172. Yang, X., Boehm, J. S., Yang, X., Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S. R., Alkan, O., Bhimdi, T. (2011). A public genome-scale lentiviral expression library of human ORFs. Nat. Methods 8, 659-661.

173. Yang, F., Sun, S., Tan, G., Costanzo, M., Hill, D. E., Vidal, M., Andrews, B. J., Boone, C. and Roth, F. P. (2017). Identifying pathogenicity of human variants via paralog-based yeast complementation. PLoS Genet. 13, e1006779.

174. Young, A. I., Benonisdottir, S., Przeworski, M. and Kong, A. (2019). Deconstructing the sources of genotype-phenotype associations in humans. Science 365, 1396-1400.

175. Zhang, Z., Li, C., Wu, F., Ma, R., Luan, J., Yang, F., Liu, W., Wang, L., Zhang, S., Liu, Y. (2015). Genomic variations of the mevalonate pathway in porokeratosis. eLife 4, e06322.

176. Zhong, Q., Pevzner, S. J., Hao, T., Wang, Y., Mosca, R., Menche, J., Taipale, M., Taş an, M., Fan, C., Yang, X. (2016). An inter-species protein-protein interaction network across vast evolutionary distance. Mol. Syst. Biol. 12, 865.