CRISPR-Cas9 Induced Combinatorial Genome Editing in Saccharomyces cerevisiae

Brittany Greco

A
Thesis
In the Department of Biology

Presented in Partial Fulfillment of the Requirements for the Degree of

Master of Science (Biology) at

Concordia University
Montreal, Quebec, Canada

December 2020

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By**: Brittany Greco**

Entitled: **CRISPR-Cas9 Induced Combinatorial Genome Editing in *Saccharomyces cerevisiae***

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

Master of Science (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. Vincent Martin

_____Examiner

Dr. Vincent Martin

_____Examiner

Dr. Alisa Piekny

_____External Examiner

Dr. Michael Hallett

_____Supervisor

Dr. Aashiq Kachroo

Approved by

_____Graduate Program Director

Dr. Robert Weladji,

_____Dean of Faculty

Dr. Pascale Sicotte

2020/2021

Abstract

CRISPR-Cas9 Induced Combinatorial Genome Editing in Saccharomyces cerevisiae

Brittany Greco

For decades, Baker's yeast (*Saccharomyces cerevisiae*) has served as a tremendous model for biomedical research. Genome-scale engineering in yeast is feasible primarily due to prodigious homology-directed DNA repair, a plethora of genetic tools/selection markers, and simple conversion between haploid and diploid forms. However, with the emergence of yeast as a model eukaryote for systems and synthetic biology research, there is a need for highly efficient and scalable genome engineering strategies. Previously, using CRISPR-Cas9, our laboratory developed a method for one-step, marker-free editing of the yeast genome using Homology Directed Repair (HDR). In the first part of my work, I created CRISPR-Cas9 toolkits targeting several yeast loci and showed specific, efficient and targeted Double-strand Breaks (DSBs) followed by HDR. Next, by combining CRISPR-Cas9, DNA repair via HDR, and yeast mating and sporulation, I demonstrate a highly efficient gene drive (referred to as Cas9-induced Gene Drive or CGD) to perform precise, scar-free, and selection-less conversion of native yeast loci to heterologous engineered loci respectively. To test the efficiency of the gene drive, I convert the functional copy of the *Ade2* gene to *Ade2* null locus using a heterozygous diploid strain (*Ade2* / Δ*ade2*::*KanMX*). First, I show the conversion of the *Ade2* to Δ*ade2*::*KanMX* locus is near 100% efficient since the DSB-resistant *KanMX* copy of the homologous chromosome serves as a highly effective repair template for HDR. Next, I demonstrate the conversion of two or more heterozygous human-yeast loci to become homozygous for human genes at a comparable rate (i.e., ~100%). To show the feasibility and scalability of this method for assembling multi-gene biosynthetic pathways or complexes, I am testing the engineering of the entire carotenoid pathway

and the orthologous human proteasome alpha core genes in yeast. Thus, CGD lays the foundation for large-scale combinatorial engineering of the entire heterologous biological processes in budding yeast.

Acknowledgements

All of this work could not have been possible without the help of so many people. Firstly, I would like to thank Dr. Aashiq Kachroo for giving me an opportunity to pursue a research project as an undergraduate student and letting it blossom as a MSc project. Thank you for believing in me and pushing me to strive and think as a scientist.

I would also like to thank my committee members Dr. Vincent Martin, Dr. Alisa Piekny and Dr. Mike Hallett for their valuable input and comments regarding this work.

A huge thank you to the *Genome Foundry* platform coordinators Dr. Nicholas Gold and Dr. Smita Amarnath for their help using the cool machines in the foundry, consequently making my life much easier. A special thank you to Dr. Smita Amarnath for her guidance and help troubleshooting experiments.

A thank you to Dr. Micheal Pyne for the engineered carotenoid pathway strains and to our collaborators in Austin, New York City and Toronto.

I'd like to also thank the members of the Kachroo lab for the amazing lab environment from the "Wall of Scientific Horrors" to the "Zen Zone" to the Halloween/ Christmas decorating competitions to starting a labelling war with the Hallett Lab. A particular thank you to Mudabir Abdullah for contributing to this work and for working on this beast of a project with me and of course for our daily chitchats and coffee breaks. Additionally, I would like to thank all the people from the other labs in GE and SP for their support and catch ups. And of course, to all my friends who have been my rocks throughout the last few years, always encouraging me to follow my dreams and reminding me to keep at it.

Lastly, I'd like to thank my dad who contributed suggestions for thesis titles. Here are a few that didn't make the cut. "CRISPR: Not Just a Light Snack", "Yeast: From Making Bread to Saving Lives", "Yeast: From the Food Network to the Discovery Channel", "Rise Up to the Challenge, Yeast: the Pathway to New and Existing Scientific Research". He tried.

Dedications

This thesis is dedicated to the Greco, Crecca, Tita, Filippelli and Kuleszka family here and in spirit. I want to take this moment to thank my mom and dad for their never ending support and motivation to keep working through the tough times but also enjoy the great times. To my brother Daniel for supporting his little sister through her academic career. To my grandparents who were prime examples that hard work will always pay off and to my aunt Susan and uncle Pat whose careers motivated me to pursue my passion. Lastly, a special dedication to my godmother Lily, who inspired me to pursue a career in research through your fighting spirit.

I am eternally grateful for all their support even though they rarely understood what my research was about but still smiled and nodded as I kept talking about it.

Finally, I want to dedicate this work to my 10 year old self who wrote a report about how she wanted to become a scientist. 15 years later, we made it girl and we are striving for greatness.

Contribution of Authors

Table 1. *Alpha* sgRNA targeting plasmids were created by Mudabir Abdullah.

Figure 12. The CRISPR-Cas9 experiments and genotyping was performed by Mudabir Abdullah.

Figure 13. Tetrad dissection, mating experiments and genotyping was done by Mudabir Abdullah.

Figure 14. Single carotenoid gene integration in yeast was done in collaboration with Dr. Michael Pyne.

Table of Contents

List of Figures

List of Tables

1. Chapter1- Background

1.1. Using CRISPR-Cas9 to induce DSBs and integrate foreign DNA in yeast

Yeast has innately high homologous recombination rates that simplify genomic manipulation [1–3]. Integration of foreign DNA in the budding yeast genome was first shown via **H**omologous **R**ecombination (HR) using linear and circular DNA bearing yeast homologies to target sites [4,5]. Additionally, having unique restriction enzymes sites flanking the homologous regions significantly increased the chances of DNA integration after the generation of the Double-Strand Breaks (DSBs) [6]. This increased integration efficacy was shown after a DSB was induced in the genome which in turn promoted integration of DNA via **H**omology-**D**irected **R**epair (HDR) [7,8]. With the combination of homologous recombination, auxotrophic selection and the induction of DSBs in genomes, meganucleases such as *HO* and I-*Sce*I were used to allow modification in yeast cells with greater efficiency than previously established methods [9–12]. Meganucleases specifically cut DNA and have large recognition sites ranging from 18-24 base pairs [12]. Having two sites of these meganucleases on both sides of the homology regions generate DSBs after the expression of these enzymes [8–11]. Though these methods of DNA modification have been successfully established in yeast, they do present some issues. This method of DNA modification while scarless, requires integration of the required restriction enzymes sites prior to DNA manipulation. Moreover, they require additional selection and confirmation for each step of integration. Overall, the methods are not suitable for scalability. We opted for another method of DNA modification using CRISPR-Cas9 which requires fewer steps

to induce a specific DSBs in yeast, needs no prior integration of foreign restriction enzymes sites and yet still allows for efficient yeast DNA manipulation.

The CRISPR-Cas9 system was first discovered in bacteria, serving as an immunity response system to specifically target and destroy invading pathogens [13]. The system is composed of a scaffold guide RNA (sgRNA) that is complementary to a specific DNA sequence. This scaffold guide RNA (sgRNA) sequence is recognized by a Cas9 protein. Together, the Cas9-sgRNA complex induces a double-stranded break (DSB) in a sequence-specific manner. While bacterial cells use this system to kill invading viruses, this technology can be used to selectively target any piece of DNA in any organism [1]. The Cas protein (Cas9) used in this study carries the HNH and RucC nuclease domains that allow for both strands of DNA to be cut [14,15]. We previously exploited the CRISPR-Cas9 system in a haploid yeast strain to perform precise genome editing in yeast [16,17]. DSBs made by the CRISPR-Cas9 system are lethal in yeast unless repaired mainly by Homology-Directed Repair (HDR) [18,19]. While yeast can repair DSBs in DNA in two different ways: Non–Homologous End Joining (NHEJ) and Homology Directed Repair (HDR), HDR tends to be favoured and more efficient when repairing a DSB in yeast [12,20,21]. Though both pathways operate in budding yeast, they function in a mutually exclusive manner, such that if one is active, the other one is suppressed [22,23]. Furthermore, NHEJ may successfully repair the DSBs in yeast without any errors, however, those loci are prone to further targeting by the Cas9-gRNA complex, thus, those yeast cells never survive. On the other hand, NHEJ via error-prone repair occurs at lower efficiency, thus, the yeast cells that survive also harbour mutations or deletions at the edited locus [20]. The HDR pathway can also repair the DSB, however, it requires a DNA template bearing a sequence homology to the affected

locus, which allows for efficient and precise DNA repair [8,19]. Previously, our laboratory established a scar-free, selection-less system to precisely edit the yeast genome. The sgRNAs homologous to yeast genes were designed to target yeast loci. The expression of the sgRNA-Cas9 system induces a double-stranded break in yeast. In the absence of and owing to low error-prone NHEJ efficiency, the majority of the yeast cells die due to the regeneration of the cut site. However, to allow the cells to repair via HDR, a repair template is provided with sequence homology at the 5' and 3' ends of the DSB.

1.2. Humanization of yeast from one gene at a time to entire systems

The most deeply evolutionarily conserved human genes encode essential cellular machinery whose failures are linked to diverse diseases [24,25]. While humans and yeast obviously differ dramatically with respect to cell and tissue organization (i.e., unicellular vs. multicellular body plans), diet, metabolism, motility, and environment, we still share several thousand protein-coding genes with yeast [17,26–31]. Despite over a billion years of evolutionary divergence, the remarkable extent to which protein-coding genes are still functionally equivalent between humans and yeast emphasizes the power even of distant organisms for studying human gene function [17,26–31]. Recent systematic studies have identified many orthologous human genes that can individually complement a lethal growth defect conferred by the loss of the corresponding yeast gene with little or no effect on growth [17,26,31]. These studies reveal a striking trend: humanizability is not well-explained by sequence similarity between the human and yeast genes. Instead it is a property of local physical or genetic interaction within protein complexes and pathways. We refer to this feature as "genetic modularity", such as some systems are entirely non-replaceable (e.g., DNA replication initiation complexes), whereas, some modules

are entirely replaceable one gene at time, including the proteasome (**Figure 1A**), sterol and heme biosynthesis pathways [17,26,31]. Moreover, some systems, such as the proteasome complex, have sub-modules that are replaceable while some are not (**Figure 1A**) [17,26,31]. However, genes do not work in isolation. They often interact genetically and physically with several other genes. One caveat of humanization assays, therefore, could be that singly humanized yeast strains aren't truly replicating the human systems. However, are higher order humanizations possible? The modularity paradigm allows us to test if the entire yeast and human systems are, to a first approximation, interchangeable (at least in yeast) (**Figure 1B**). However, the route to full humanization of entire pathways or complexes in yeast is unknown. Therefore, we propose a method (C-Gene Drive or CGD) that allows us to use singly engineered humanized strains to

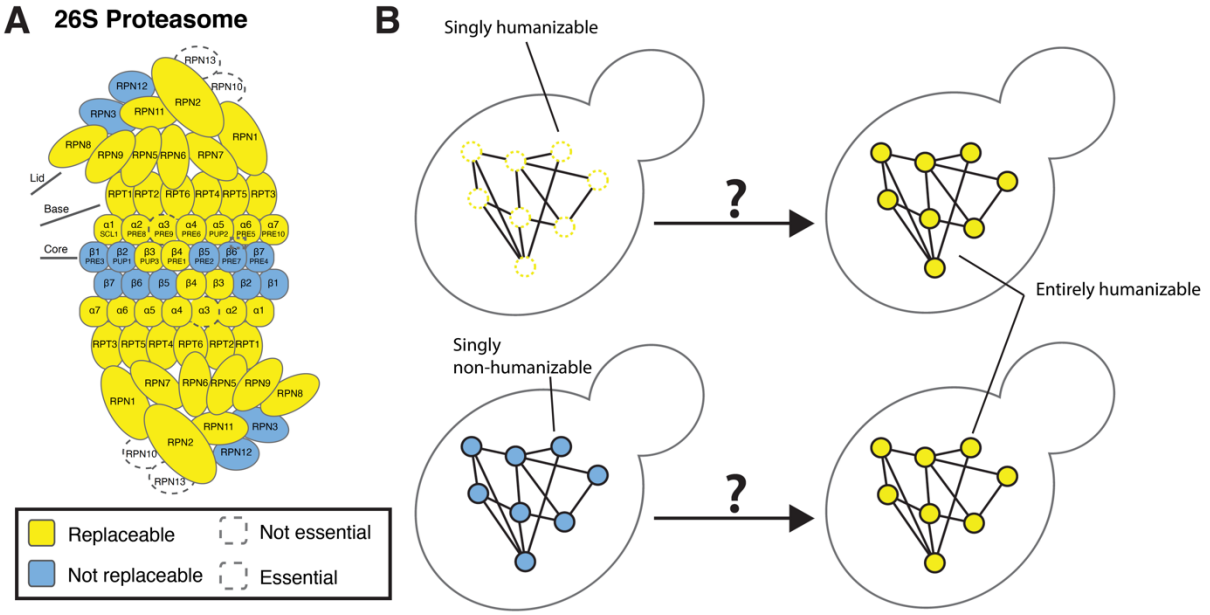


Figure 1. From humanizing single genes to entire cellular processes in yeast.

A. Plasmid-based complementation assays reveal that the yeast proteasome complex is largely replaceable by their human counterparts one-gene-at-a-time. However, certain sub-modules such as the subunits in the beta-core and lid aren't replaceable. **B.** Biological processes in livings cells are often carried out by a multitude of genes forming a network as indicated by circles connected via lines showing genetic or physical interactions. Singly replaced yeast genes by their human equivalents that belong to the same biological process suggest that the system could be humanizable in their entirety. This hypothesis may be true for the systems that are not replaceable even as single genes but may become replaceable as entire systems.

build fully engineered systems in a single yeast strain. We also propose to use the method to evaluate the possibility of building systems and modules that are individually non-replaceable **(Figure 1B**). While we intend to use the CGD method to perform higher order humanizations, the method is generally applicable to build any heterologous system in yeast. To show the efficiency and the general use of the CGD method, we will use a colored colony read out , such as yeast *ade2* mutants that appear as red colonies and yeast with heterologous carotenoid pathway that appear as red or orange colonies depending on the partial (3 gene) or complete (4 gene) assembly respectively. For the humanization of entire yeast systems, we will focus on the humanization of the entire proteasome core and base modules to build a synthetic platform for testing human genetic variation and drug discovery in a simplified cell.

1.3. Engineering entire biological processes in yeast - a challenge

To build a fully humanized or heterologous system in a yeast surrogate, we need a highly scalable and efficient method for precise combination and conversion of engineered yeast loci. In fact, yeast has a rich history of combining genetically engineered loci such as in the case of testing gene-gene interactions [32,33]. The methodology, also referred to as Synthetic Genetic Array (SGA) uses mating and sporulation to combine genetic loci (such as gene knockouts) by using selectable markers to select for haploid cells after sporulation [33]. However, when >2 loci are involved, the strategy is limited by the availability of selection markers. A strategy referred to as "GFP monster" tested the use of GFP expression as a readout of a combination of engineered loci wherein a GFP expression cassette replaces the yeast locus [34] . The more the number of GFPs in a strain, the higher the level of fluorescence that can be measured by using FACS. While this strategy is useful for combining deleted loci, it will not be useful for humanization or heterologous

pathway engineering which requires the insertion of a foreign gene replacing the yeast locus. Alternate strategies such as plasmid shuffling and the use of YACs are useful. These strategies assemble the entire biological system in a single plasmid. However, even those methods require additional steps that involve sequential deletion of the native systems or pathways [35–37]. Plasmid-based complementation is also restricted to the number of genes that can be cloned into one plasmid and YACs require a lot of complex cloning that may be time consuming. Alternatively, the sequential engineering of systems is possible; however, it gets increasingly difficult as the number of loci to be engineered increases (**Figure 2**). The sequential assembly is also a painful guessing game as some partially assembled intermediates may be less fit (**Figure 2**). For example, for a simple 3-gene system, one requires at least 9 intermediate genotypes. Furthermore, the fitness landscape of the intermediate hybrid strains is unknown, thus, making the process a guessing game to identify a viable route for the assembly of the entire system (**Figure 2**).

In order to bypass these issues, we propose using a combination of our previously established CRISPR-Cas9 based method [16,26], a subset of previously individually humanized strains and a series of mating/sporulation cycles to allow cells to naturally and directionally (fitness-driven) build engineered biological processes in yeast. Mating allows cells to innately combine engineered loci with wildtype loci and furthermore combine two or more engineered loci. Sporulation permits the independent segregation of chromosomes which together facilitates combination and conversion of loci by a creating strain with the desired genotype (>1 engineered locus). At a diploid stage, the CRISPR-Cas9 allows us to select for specific genotypes of combined loci since one allele will carry a wildtype yeast gene and the other will carry an

engineered CRISPR-Cas9 resistant gene. Next stage involves the elimination of a specific yeast locus from a yeast population mimicking a "**gene-drive**" strategy by using CRISPR-Cas9-sgRNA targeting the yeast locus. This is possible because when a DSB is induced at a particular locus, the engineered locus flanked by the entire homologous chromosome will serve as a highly stable

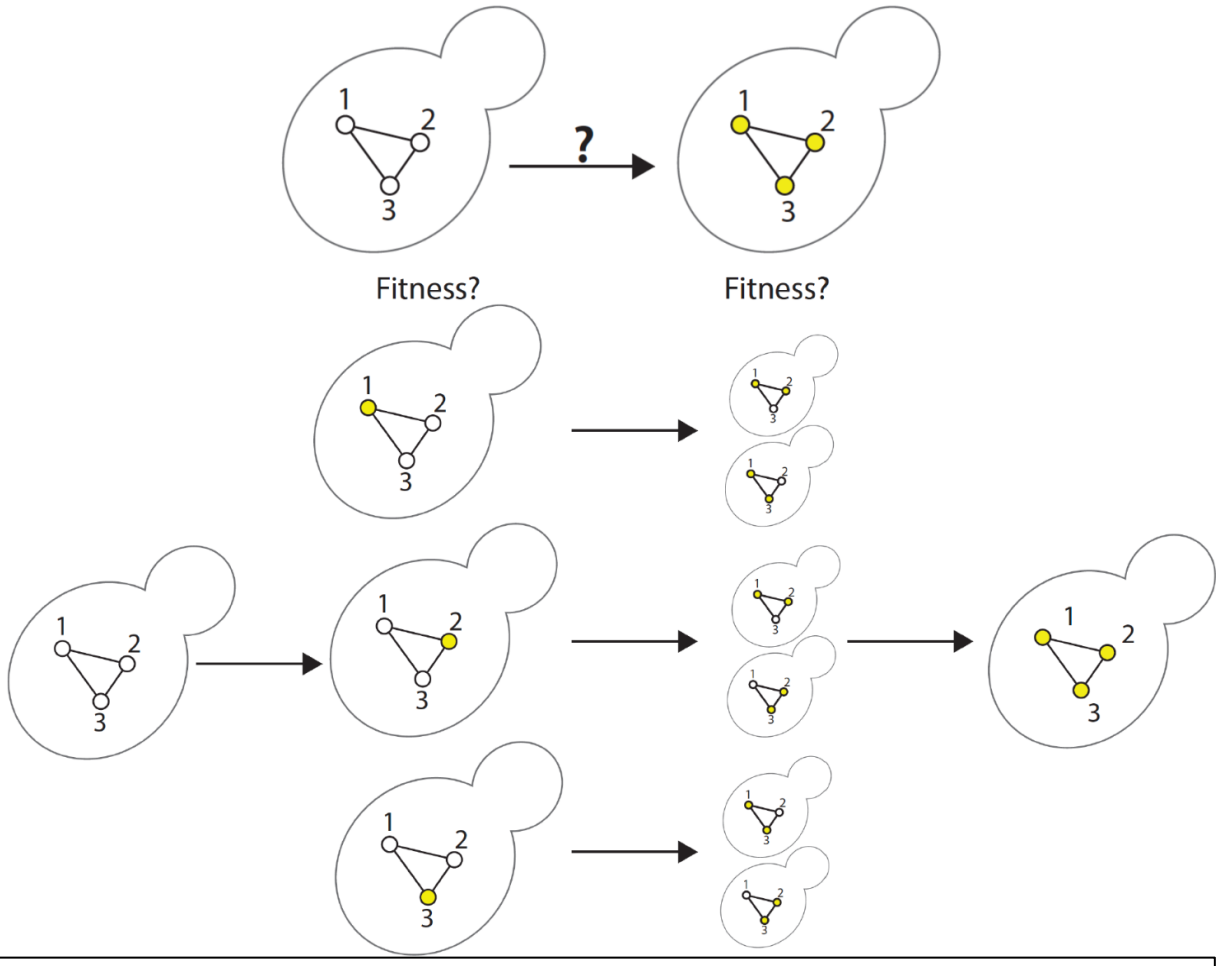


Figure 2. Multiple routes to engineering of biological systems using a sequential replacement strategy.

Let's examine a simple 3 gene pathway. How do we go from 1 engineered locus to all engineered loci in single strain? If gene 1 is humanized first, then a choice is made to humanize gene 2 or gene 3 and then humanization of the last gene proceeds; that is if these intermediates are viable or not. The same process would occur if humanization started with gene 2 or gene 3. This sequential replacement strategy would require 9 possible intermediate genotypes before achieving a fully engineered system. These intermediate genotypes would be far higher as the scale of genetic systems gets complex. This shows that sequential humanization which was once used to build pathway is time-consuming and does not address fitness levels at each intermediate.

and efficient repair template to create homozygous diploids. We propose to use this method to combine, convert and select any locus of choice in yeast. In this work, I will show how this strategy works by testing the efficiency in the adenine pathway (*Ade2* locus serves as an easy readout of locus conversion*)*, followed by testing some combinations of essential genes in the proteasome core/base. Finally, I apply the CGD method to build a heterologous pathway, such as the carotenoid pathway, that provides a colony colour phenotype as a proxy for pathway assembly using our established C-Gene Drive method.

To summarise, I am developing a new, next-generation method to combinatorically engineer heterologous systems in yeast. While we aim to engineer several critical human systems using this strategy, the method is equally applicable in combining any engineered loci since the only requirement is the CRISPR-Cas9-sgRNA mediated DSB.

1.4. The problem addressed

The advancement in *de novo* DNA synthesis, next-gen sequencing technologies and a greater understanding of diverse biochemistry across the tree of life, a new era of biology is dawning - synthetic genomics and synthetic biology. Genetic engineering a single gene to entire biosynthetic pathways or protein complexes is becoming possible. Simple and well-studied cells like *E. coli* (a prokaryote) and Yeast (a eukaryote) serve as biological test tubes providing an excellent bioengineering platform owing to their staggering genetic tractability, versatile metabolism, and ease of culture in the lab. Several decades of fundamental research has established yeast as an ideal model eukaryote for system genetics and synthetic biology [38]. However, to build complex heterologous multi-gene biosynthetic pathways or protein complexes

in yeast on a massive scale requires new genetic tools. Herein, we present the fundamentals of one such technology and provide a blueprint for scalability.

For decades now, yeast has served as a tremendous model for studying basic eukaryotic biology. Yeast have innately high homologous recombination efficiency that simplifies genomic manipulation [1–3]. The ease of genetic manipulation has provided several resources such as genome-wide knockout, temperature-sensitive, heterologous regulatable promoter, etc., collections of yeast [39–41]. Additionally, the availability of many selectable genetic markers and simple conversion between haploid and diploid forms has provided avenues to easily combine genetically engineered loci to understand gene-gene interactions at a global scale [27,32,33]. However, for multi-gene engineering of heterologous pathways in yeast, we will either require several new selection methods or be able to selectively remove markers to be used reiteratively [42]. Alternatively, the engineered heterologous pathways can be built on plasmids or yeast artificial chromosomes (YACs) [35,37]. However, if the expression of heterologous pathways also tests the functional replaceability of the native yeast processes, it further involves several sequential steps of knocking out the individual orthologous yeast genes [35,37].

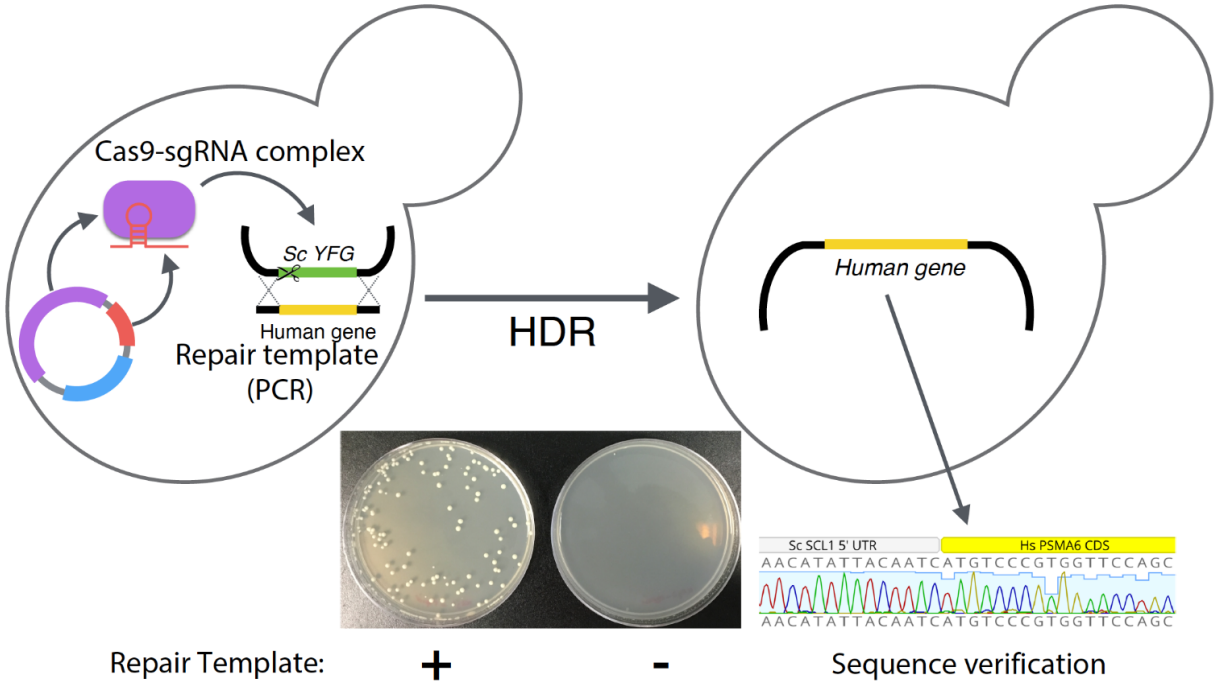


Figure 3 .CRISPR-Cas9 based genome editing in yeast.

CRISPR-Cas9 mediated DSB is lethal to most yeast cells. However, DSBs are efficiently repaired via HDR using a repair template that also provides resistance to cutting by CRISPR-Cas9-sgRNA. As an example, shown here, a repair template can be a PCR amplified human gene with a flanking yeast specific homologous sequence. Functionally replacement of the yeast gene by a corresponding human gene allows the cells to survive. The survivors often harbor precise replacement of the human gene as identified by Sanger sequencing.

Recently, our laboratory demonstrated the use of CRISPR-Cas9 for rapid replacement of yeast loci (**Figure 3)** [16,26]. CRISPR-Cas9-mediated double-strand breaks can be applied repeatedly for sequential replacement of multiple loci within the same strain. While this is a straightforward approach, it may not be the most effective strategy for a very large multi-gene system. Moreover, it is not easy to perform multiple replacements in parallel. When attempting to replace a locus, all other remaining loci must wait until that strain is constructed (or abandoned if not feasible). Yeast mating allows us to adapt the CRISPR-Cas9 editing method and have a way of easily combining replacements.

In this work, by using CRISPR-Cas9 mediated genetic manipulation of yeast loci, I show a scar-less and selection-less precision conversion/combination of many yeast loci. The method, therefore, circumvents the need of selection markers while also allowing the engineering of heterologous pathways at scale.

2. Chapter 2- A Resource Of Many CRISPR-Cas9-sgRNA Reagents For Budding Yeast

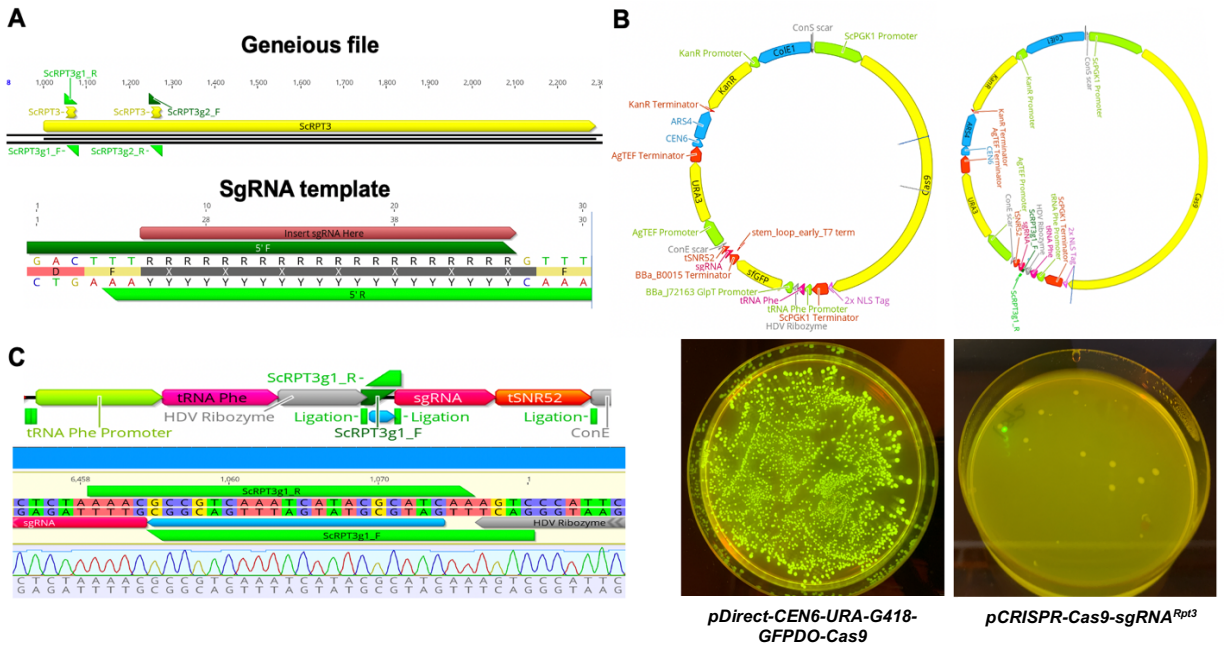
2.1. Direct cloning of many yeast-specific CRISPR-Cas9-sgRNA vectors

Previously, our laboratory established a CRISPR-Cas9 based strategy to target both essential and non-essential yeast genes [16,26]. The efficiency of precise genetic manipulations in yeast is largely possible due to significantly lower error-prone NHEJ mediated DSB repair in yeast as compared to HDR [20]. The CRISPR-Cas9 method is also efficient in engineering heterologous genes at safe harbor sites within the yeast genome. These non-coding sites are scattered in the yeast genome and allow for a foreign DNA integration with negligible fitness cost to the host cell [43].

The generation of CRISPR-Cas9-sgRNA reagents require specific assembly of DNA to express a sgRNA that will allow for specific locus targeting and a Cas9 expression cassette that together after translation can initiate a double-stranded break at said location. Towards that , we designed a golden gate compatible shuttle vector for Cas9-sgRNA expression using the *MoClo* collection [44]. This shuttle vector is capable of expressing both a sgRNA and a Cas9 protein in yeast. Golden Gate cloning is a type of cloning method that uses type II restriction enzymes that cut outside of from their recognition sites that allows for the directional assembly of multiple parts in a single or sequential steps [16,26,44].

The sgRNAs were designed using the Geneious software that uses algorithms based on Doench *et al.* to selectively identify and score CRISPR sites within the yeast genome [45] . The software scores these target sites by scanning the whole yeast and look for sequencing similarity [45,46]. We designed at least 2 CRISPR sites per locus preferably located near the beginning of

the gene and picked the ones with low OFF-target and high ON-target activity for low background. We hypothesize that inducing a DSB in the beginning of a gene will ensure a non-functional yeast protein when an error-prone NHEJ pathway tends to repair the locus independent of HDR [20]. Using the golden gate cloning strategy [16,26,44], we swapped out the *GFP* for a sgRNA cassette (in the form of two complementary primers with unique 4-base overhangs). The 20 base CRISPR sgRNA sequence was copied into a Geneious sgRNA template with golden gate compatible 5' and 3' overhangs respectively (**Figure 4A**). The annealed forward and reverse primers generated from this sgRNA template allowed for a one-step cloning into the **p*Direct-***



pDirect-CEN6-URA-G418-GFPDO-Cas9

pCRISPR-Cas9-sgRNA^{Rpt3}

Figure 4. Building single CRISPR-Cas9-sgRNA expression vectors *in silico* & *in vitro*.

A. Geneious file of a gene of interest (*Rpt3* from yeast proteasome base - TOP panel), shows two CRISPR sites located at the beginning of the gene that have a high ON-target and low OFF-target scores. These CRISPR sites are inserted into an sgRNA template). The primers are synthesized such that when annealed the sgRNA templates (BOTTOM panel) create 5' and 3' golden gate compatible 4 base overhangs that allow cloning into the **p*Direct-CEN6-URA-G418-GFPDO-Cas9*** vector. **B.** The plasmid map of **p*Direct-CEN6-URA-G418-GFPDO-Cas9*** (TOP panel) when transformed in *E. coli* results in bright green colonies (BOTTOM panel). However, the insertion of sgRNA cassette results in swapping out of the *GFP* as indicated in largely non-fluorescent *E. coli* colonies (BOTTOM panel). **C.** Sequence verification of the sgRNA region of the clone (**pCRISPR-Cas9-sgRNA^{Rpt3}**) shows correct sequence alignment with the reference sequence.

CEN6-URA-G418-GFPDO-Cas9 plasmid and transformed into *E. coli.* The *GFP* dropout allowed for colony screening for the clone of interest as *GFP* negative colonies likely contain sgRNA cassette insert (**Figure 4B**). These *GFP* negative colonies were picked to isolate the plasmid followed by verification using restriction enzyme digestion and Sanger sequencing (**Figure 4C**).

Table 1. CRISPR site sequences cloned for sgRNA expression in yeast

Locus	sgRNA Name	CRISPR Sequence (sequence 5'→ 3')
Ade2	*Ade2*-sgRNA1	AATTGTAGAGACTATCCACAAGG
Ade2	*Ade2*-sgRNA2	CTGCTCATAGAACTCCACATAGG
Rpt1	*Rpt1*-sgRNA1	ATTCAAGTCTTAAAATCATACGG
Rpt1	*Rpt1*-sgRNA2	GCGAGAATCAAAGAAAAGGCCGG
Rpt2	*Rpt2*-sgRNA1	ATTCAAGTCTTAAAATCATACGG
Rpt2	*Rpt2*-sgRNA2	CCAAATTTAGATTGAACGGGAGG
Rpt3	*Rpt3*-sgRNA1	TCCTCCTGCTTCTTTTCAAAGGG
Rpt3	*Rpt3*-sgRNA2	AGAAGTAAAGAGAATCCAGTCGG
Rpt4	*Rpt4*-sgRNA1	AGAAGTAAAGAGAATCCAGTCGG
Rpt4	*Rpt4*-sgRNA2	CCCCCAATCCTGCTAATAAAGGG
Rpt5	*Rpt5*-sgRNA1	CTGTTGGGTGTGATTGTCGCCGG
Rpt5	*Rpt5*-sgRNA2	ACAACGTTAGCCACAAGGTACGG
Rpt6	*Rpt6*-sgRNA1	GTATTAGAAACCCACGAAAGTGG
Rpt6	*Rpt6*-sgRNA2	GTATTAGAAACCCACGAAAGTGG
Alpha3	*Alpha*3sgRNA1	AGATGGACACATTTTCCAAGNGG
Alpha4	*Alpha4*-sgRNA 1	TTGGGATTATGGCATCTGATNGG
Alpha7	*Alpha7*-sgRNA1	ATCGGTATAAAGTGTAACGANGG
511b	*511b*-sgRNA1	CCGTGGCTGACTGGCATACACTG
USERX-1	*USERX-1*-sgRNA1	GTAGCTACAAGAACATATGGTGG
*FgF2*0	*FgF2*0-sgRNA1	GTTAGAGCTGTTACAAGTTACGG
FgF24	*FgF24*-sgRNA1	CCTATTGGACAAGATTTACGAGG

The sgRNAs in **Table 1** represent the yeast genes I worked on during my Master's research project. The sgRNAs target the base of the proteasome that is made up of six *Rp*t genes (*Rpt1 to 6*) [47,48]. The proteasome core is made up seven *Alpha* and *Beta* subunits though only three (*Alpha3,4,7)* were studied in this work [49]. I also used the *Ade2* gene in the adenine synthesis pathway as a colony colour readout of efficiency of DNA repair via NHEJ or HDR. The mutated *ade2* locus results in red coloured colonies. Lastly, I used various non-coding, safe-harbour sites in the yeast genome (*511b, USERX-1, FgF20, FgF24*) to integrate the genes from the carotenoid pathway that also provide a colony colour readout as a proxy for pathway engineering [50–52].

Once confirmed, each Cas9 & sgRNA expression vector was tested in a haploid wildtype yeast strain (BY4741). As an example, we show the results for a representative set of essential yeast genes and a non-essential yeast gene, *Ade2,* that when mutated, results in a red colony phenotype [53]. Briefly, in the case of the mutated *Ade2* yeast locus, the lack of conversion of p-ribosyl aminoimidazole (substrate for *Ade2*) to p-ribosyl imidazolecarboxylate causes the accumulation of substrate (p-ribosyl aminoimidazole) - a red colored pigment that makes the colonies appear red [53].

Our first test of whether pCRISPR-Cas9-sgRNA^{gene} is functional in yeast is to observe if the plasmid transformation is lethal (Observed Colony Forming Units or CFU_O). As a control, we transform the same yeast cell with the plasmid that only expresses Cas9 but not the sgRNA. The control transformations provide an estimate of the efficiency of the plasmid uptake by yeast cells (Expected Colony Forming Units or CFU_E). Typically, we use 500ng to 1µg plasmid for each transformation which results in ~10^3 - 10^4 CFUs on a selection plate (**Figure 5, Top panel**). The efficiency of the CRISPR cut is, therefore, calculated as using a simple formula of CFU_O/CFU_E with the value close to zero suggesting high efficiency. For example, upon transformation of

pCRISPR-Cas9-sgRNA1^{Ade2}, and due to a DSB at the *Ade2* locus, the number of colonies on a selection medium is drastically reduced (0-20 colonies on average, **Figure 5, Bottom left panel**) compared to the empty vector control transformations (~10^3 colonies on average, **Figure 5, Top left panel**). This yields an efficiency of CFU_O/CFU_E, i.e. 20/1000 = 0.02 suggesting highly efficient activity (**Figure 5 inset**). Notice that all the surviving colonies show red color phenotype suggesting ON-target activity (**Figure 5, Bottom left panel)**. I have consistently observed this decrease in the number of colonies after transformation with pCRISPR-Cas9-sgRNA^{gene} irrespective of the essentiality of the locus (see **Figure 5, Bottom panel**). In the case of the CRISPR-Cas9-sgRNA targeting an essential yeast gene, the surviving colonies are far lower compared to the non-essential gene (i.e., *Ade2* locus). Together, these observations suggest that the error-prone repair of a DSB (via NHEJ) in a haploid yeast cell is less efficient as previously observed [20,21]. The lethality of the DSB induced by the pCRISPR-Cas9-sgRNA^{gene} system, therefore, serves as an indirect estimate of the efficiency of the sgRNA target sequence recognition by Cas9. As shown below, majority of our Cas9-gRNA expression plasmids performed efficiently whereas only one with high ON-target and low OFF-target scores performed poorly in our *in vivo* experiments (see **Figure 5, Bottom panel for pCRISPR-Cas9-sgRNA2^{Rpt4}**). Additionally, in the case of another pCRISPR-Cas9-sgRNA2^{Ade2}, while I observed lethality, the surviving colonies do not show red color phenotype as seen in the case of pCRISPR-Cas9-sgRNA1^{Ade2}. This behavior could be attributed to the fact that the target sequence of this sgRNA is located towards the end of the gene, which even after an error-prone NHEJ repair still makes a functional *Ade2* protein.

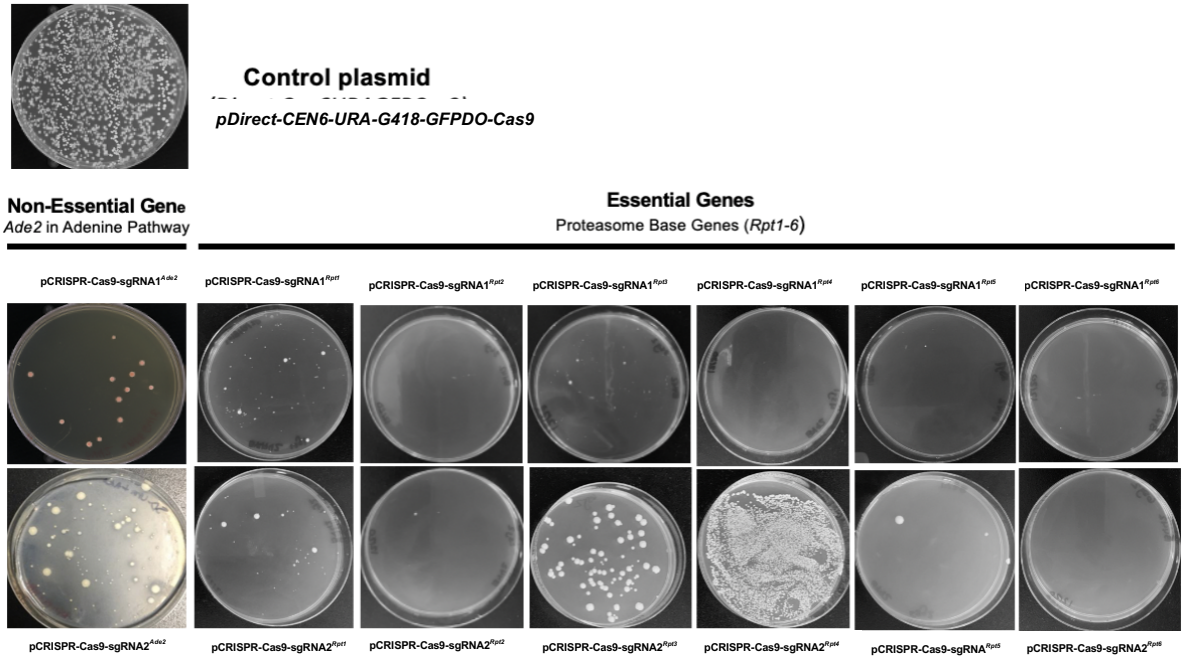


Figure 5. A CRISPR-Cas9-sgRNA induced DSB is lethal to a haploid yeast cell.

A haploid wildtype yeast strain transformed with a control plasmid harbouring only a Cas9 cassette without sgRNA sequence results in $>10^3$ CFUs (Top panel). However, the cells transformed with the plasmid carrying both Cas9 and a specific sgRNA targeting either non-essential or essential genes show a significant decrease in CFU (typically 0-20 per replicate) (**Bottom panel**). For example, cells harboring plasmid with pCRISPR-Cas9sgRNA1^{Ade2} (targeting the middle of the *Ade2* gene) showed severe lethality (only 12 CFU) and all surviving colonies show red colour phenotype suggesting the loss-of-function mutations in the gene. Comparatively, while we observe similar lethality in the case of pCRISPR-Cas9sgRNA2^{Ade2} (targeting the end of the *Ade2* gene) but the surviving cells appear as white colonies suggesting mutations at the end still allow the production of a functional protein. Similarly, two pCRISPR-Cas9-sgRNA expressions cassettes were tested for each proteasome base genes. All, except one (pCRISPR-Cas9-sgRNA2^{Rpt4}), Cas9-gRNA plasmids result in yeast lethality.

These Cas9-sgRNA expression plasmids are a part of the larger collection of sequence-verified and tested pCRISPR-Cas9-sgRNA^{gene} vectors (referred to as ***Sc*EDIT**). The ***Sc*EDIT** constitutes plasmids targeting several non-essential yeast loci belonging to the adenine and glycosylation pathway, certain safe-harbor sites (collaboration with Drs. M. Pyne and V. Martin) for engineering heterologous pathways such as the carotenoid pathway. The collection also includes targets for a many essential genes such as the proteasome complex core and base, the

sterol and heme biosynthesis pathway, the yeast cytoskeleton and some tRNA genes. Additionally, ***Sc*EDIT** also includes vectors for the expression of >2 sgRNA transcription units simultaneously.

For each of the pCRISPR-Cas9-sgRNA^{gene}, we have also tested the ON-target activity by specifically inserting a repair template (via HDR) at the cut locus as described in **Chapter 3**.

3. Chapter 3- Using CRISPR-Cas9 to engineer single yeast loci

3.1. Replacing yeast genes with their corresponding human orthologues at their native yeast loci

Majority of the previous humanization assays were performed using plasmid-based functional complementation assays [17,26]. In order to generate humanized yeast controlled for strain background and native expression, we propose to replace the human genes at their native yeast loci. Towards that, the yeast pCRISPR-Cas9-sgRNA^{gene} expression vectors created were tested for their efficiency at creating a specific DSB, leading to lethality in a haploid yeast cell (**Chapter 2**). We showed that the co-expression of Cas9 & sgRNA kills the majority of yeast cells with a 1-10 colonies surviving due to reduced and error-prone NHEJ. However, by co-transforming Cas9-sgRNA plasmid along with the a repair template with sequence homology to the locus, yeast survive an otherwise lethal DSB using the HDR DNA repair pathway. This strategy allows for an efficient and scar-less replacement of yeast genes with their corresponding human counterparts. I employed the following strategy for the repair template design (**Figure 6**). Hybrid primers bearing ~20 bp sequence complementarity (indicated in yellow) to the human gene and ~60-80bp homology to the 5' and 3' UnTranslated Regions (UTRs) of yeast locus (indicated in grey) were designed to build a repair template via PCR (**Figure 6**).

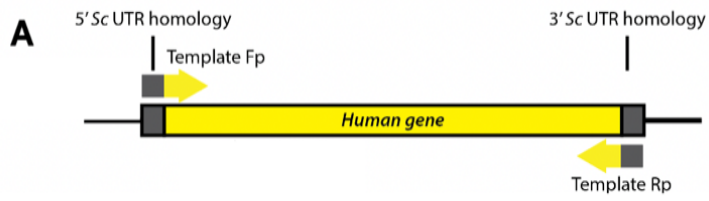


Figure 6. Generation of a human gene repair template for replacing the corresponding yeast genes at their native loci.

Schematic of human gene template showing appropriate homology and primers used for PCR amplification. Hybrid primers were used to generate a repair template via PCR. The primers harbour ~ 60-80bp homology to 5' and 3' yeast UTRs (indicated in grey) and ~20bp homology to the 5' and 3' region of the human cDNA (indicated in yellow).

Human sequence-verified cDNAs were amplified from plasmids obtained from the Human ORFeome Collection or the Mammalian Gene Collection (MGC clones) [54,55]. In this report, we tested 13 human genes for their ability to functionally replace their corresponding yeast equivalents at their native yeast loci **(Table 2).** The human genes tested for replaceability in haploid yeast cells comprise the proteasome complex specifically in the base (*Rpt* genes*)* and core *(Alpha* genes*)*.

Table 2. The list of yeast genes and their human equivalents used for integration of human genes at their native yeast loci.

Gene Function	Yeast Gene	Human Orthologue
Proteasome Base	*Rpt1*	*PSMC2*
Proteasome Base	*Rpt2*	*PSMC1*
Proteasome Base	*Rpt3*	*PSMC4*
Proteasome Base	*Rpt4*	*PSMC6*
Proteasome Base	*Rpt3*	*PSMC3*
Proteasome Base	*Rpt6*	*PSMC5*
Proteasome Core	*Alpha1 (SCL1)*	*PSMA6*
Proteasome Core	*Alpha2 (PRE8)*	*PSMA2*
Proteasome Core	*Alpha3 (PRE9)*	*PSMA4*
Proteasome Core	*Alpha4 (PRE6)*	*PSMA7 & PSMA8*
Proteasome Core	*Alpha5 (PUP2)*	*PSMA5*
Proteasome Core	*Alpha6 (PRE5)*	*PSMA1*
Proteasome Core	*Alpha7 (PRE10)*	*PSMA3*

To confirm the successful integration of human genes at their native yeast loci, I used locus specific PCR confirmations. Confirmation primers comprise a forward primer binding 150bp upstream of 5' homologous sequence used for HDR and a reverse primer binding internally to either the yeast or the human gene (**Figure 7A**). Typically, yeast gene specific primers will amplify a ~300-bp region whereas human gene specific primers will amplify ~325-500-bp region. The amplification of human gene specific PCR and the simultaneous absence of the yeast gene specific PCR suggested correct insertion of the human copy of the gene at the native yeast locus (**Figure 7B**). In certain cases, PCR amplification showed signals for both a yeast and human gene likely due to picking a polyclonal colony with mixed genotypes. These colonies were discarded.

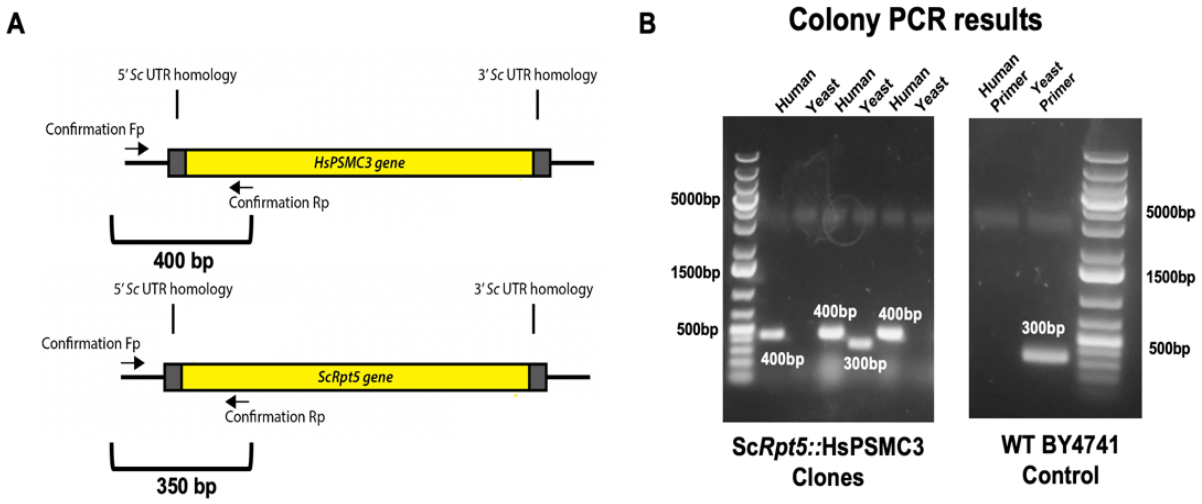


Figure 7. Confirmation of human gene integrations via colony PCR genotyping.
A. Primer design for confirming the presence or absence of human or yeast genes includes forward primers that bind 150bp upstream of the 5' homology used for HDR whereas reverse primers either bind internally to the human or yeast gene preferably amplifying PCRs of different sizes. **B.** A representative example of a colony PCR shows a clone subjected to PCR confirmation using both sets of primers (Human and Yeast) to verify successful integration. The correct clones provide PCR amplification only with human specific primers (400bp band) whereas yeast specific primers show a PCR signal only in the case of the wildtype yeast locus (300bp band).

Using the above-mentioned methodology, I successfully obtained singly humanized haploid yeast strains harboring human genes at native yeast loci for Sc*Rpt3,* Sc*Rpt5* and Sc*Rpt6* (**Figure 8**). Whereas, I haven't yet obtained human gene integration in the case of three proteasome base genes (Sc*Rpt1,* Sc*Rpt2,* Sc*Rpt4)*. Troubleshooting the lack of humanization for these yeast loci is in progress. The failure of integrations could be attributed to the low concentration of template or a need for an increased 5' and 3' homology (preferably up to 500bp). I am currently testing new transformations with upward of five micrograms of repair template with increased 5' and 3' sequence homology to the yeast locus.

After PCR confirmation, all humanized strains were sequence verified using primers outside of the repair homology and confirmed to show proper sequence alignment when mapped to the reference *in silico* humanized gene sequence (**Figure 8C**). Finally, the fitness of engineered strains was assessed using quantitative growth assays in rich YPD media at 30ºC over a period of 48 hours. The Sc*Rpt3::*Hs*PSMC4* strain grew similar to the haploid wildtype BY4741 strain whereas the Sc*Rpt5::*Hs*PSMC3* and Sc*Rpt6::*Hs*PSMC5* strains show a delay in the start of their exponential phase (**Figure 8B**). In the case of the Sc*Rpt6::*Hs*PSMC5* strain growth also plateaued at a lower OD than the rest of the strains (**Figure 8B**). These growth assays reveal modest fitness defects in humanized strains. We are in the process of the downstream biochemical characterization of the activity of the humanized yeast proteasome strains.

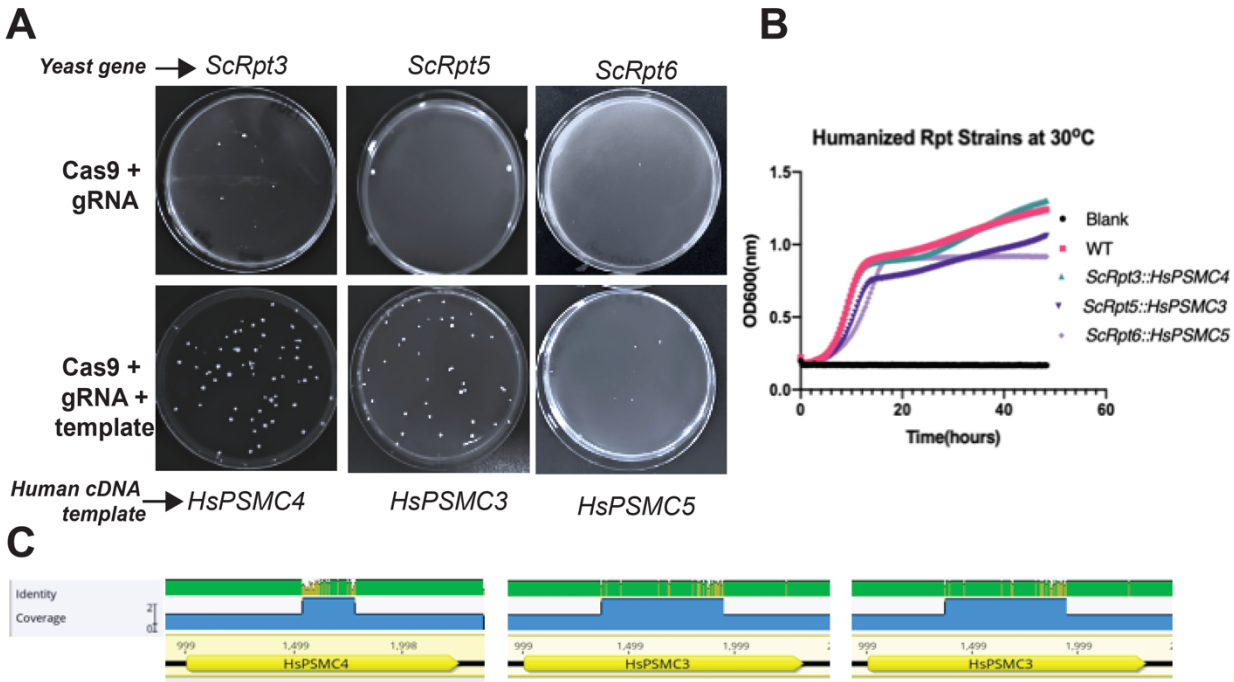


Figure 8. Humanization of individual yeast genes at their native yeast loci.

A. A haploid wildtype yeast transformed with a pCRISPR-Cas9-sgRNA plasmid targeting the Sc*Rpt3,* Sc*Rpt5,* Sc*Rpt6* yeast genes respectively leads to a lethal phenotype (Top panel). Whereas yeast co-transformed with pCRISPR-Cas9-sgRNA and a human gene (Hs*PSMC4*, Hs*PSMC*3, Hs*PSMC6* respectively) repair templates show significantly higher numbers of surviving colonies. Cells that undergo HDR when a template is available for repair show more CFUs compared to the ones that undergo DSBR via NHEJ (Bottom compared to the Top panel). **B.** Quantitative growth assays of humanized strains (*ScRpt3*::Hs*PSMC4*, Sc*Rpt5*::Hs*PSMC3*, Sc*Rpt6*::Hs*PSMC5*) show comparable growth profile compared to the wildtype strain. **D.** Sequence alignment of humanized strains confirms the humanization of the corresponding yeast loci. Sanger sequencing data was mapped to the reference humanized yeast loci using the Geneious software.

4. Chapter 4- CRISPR-Cas9-mediated Gene Drive (CGD): A Method To Perform Combinatorial Genome Editing In Yeast

4.1. A CRISPR-Cas9- mediated gene drive method to combine/convert genetically engineered loci

While we have successfully humanized yeast strains one-gene-at-a-time at their native yeast loci (**as shown in Chapter 3**). We are yet to humanized the entire biological process in the yeast context. This is a challenge especially when a combinatorial scar-less, selection-less, and multi-site genome editing in yeast is considered. In this chapter, we lay a foundation of one such method. Our proposed method, referred to as CRISPR-Cas9-mediated Gene Drive (or simply **CGD**), involves multi-level genome editing, each increasing in genotype complexity as depicted in **Figure 9**. In the following body of work, we show the preliminary results for CGD level 0, level 1 and lay the foundations for level 4.

Gene drive **level 0**, requires any singly engineered/humanized haploid yeast strain followed by mating with an opposite mating type wildtype (WT) strain, thus, generating a heterozygous diploid for that particular locus. Next, a CRISPR-Cas9-sgRNA plasmid targets a yeast sequence of the heterozygous engineered locus. The expression of Cas9-sgRNA creates a DSB only at the yeast locus while the homologous chromosome with engineered locus remains intact. The chromosome, owing to mega-base equivalent of homologous sequence serves as a stable and ideal repair template for the DSB containing chromosome. The efficient conversion of the locus from yeast to engineered one results in a homozygous diploid strain. In principle, this strategy should work at high efficiency yielding desired genotypes since any wildtype yeast loci containing strains are not allowed to propagate due to continuous targeting by Cas9-sgRNA complex. Homozygous

diploid strains can then be sporulated to obtain engineered strains of both mating types to be used for subsequent CGD levels.

In the case of CGD **level 1**, the same strategy is used to target a yeast strain harbouring >1 engineered loci by mating with the opposite mating-type wild type strain. The heterozygous diploid generated after mating is transformed with Cas9-sgRNA plasmids harbouring >1 sgRNAs expression cassettes targeting >1 heterozygous yeast loci. This strategy should only select for diploids that have converted all the yeast loci to the engineered loci. The simultaneous expression and assembly of all possible Cas9-sgRNA complexes in a strain being the only limiting factor that could prevent the conversion of loci. The sporulation of the resulting diploid strain would provide both mating type haploid engineered strains (for all the loci). The only difference between the **level 1 & 2** is that in the case of **level 2**, the engineered strains are provided by opposite mating type strains, thus resulting in combination in addition to the conversion of the engineered loci.

In the case of **level 3**, several singly engineered haploid yeast strains of both mating types are mated directionally (i.e. both the genotypes are known) followed by the transformation with an appropriate Cas9-double-sgRNA expressing plasmid respectively. This strategy should provide doubly-engineered strains of all possible combinations. For example, as shown in **Figure 9**, seven individually humanized alpha proteasome yeast strains should provide twenty-one possible combined genotypes.

In CGD **level 4**, several singly engineered strains can similarly be processed, i.e., mated, CRISPRed, and sporulated repeatedly allowing natural, random and unbiased combination of engineered loci (**Figure 9**). The survival of the strains is largely driven by the fitness of the combined engineered strains. Since the mating experiment is performed randomly, each unique

combination can be selected by the appropriate Cas9-double-sgRNA targeting only the two yeast loci respectively. Next, this one-pot reaction mix undergoes multiple rounds of mating and sporulating. At the end of each cycle, the idea is to select for unique combined engineered strains

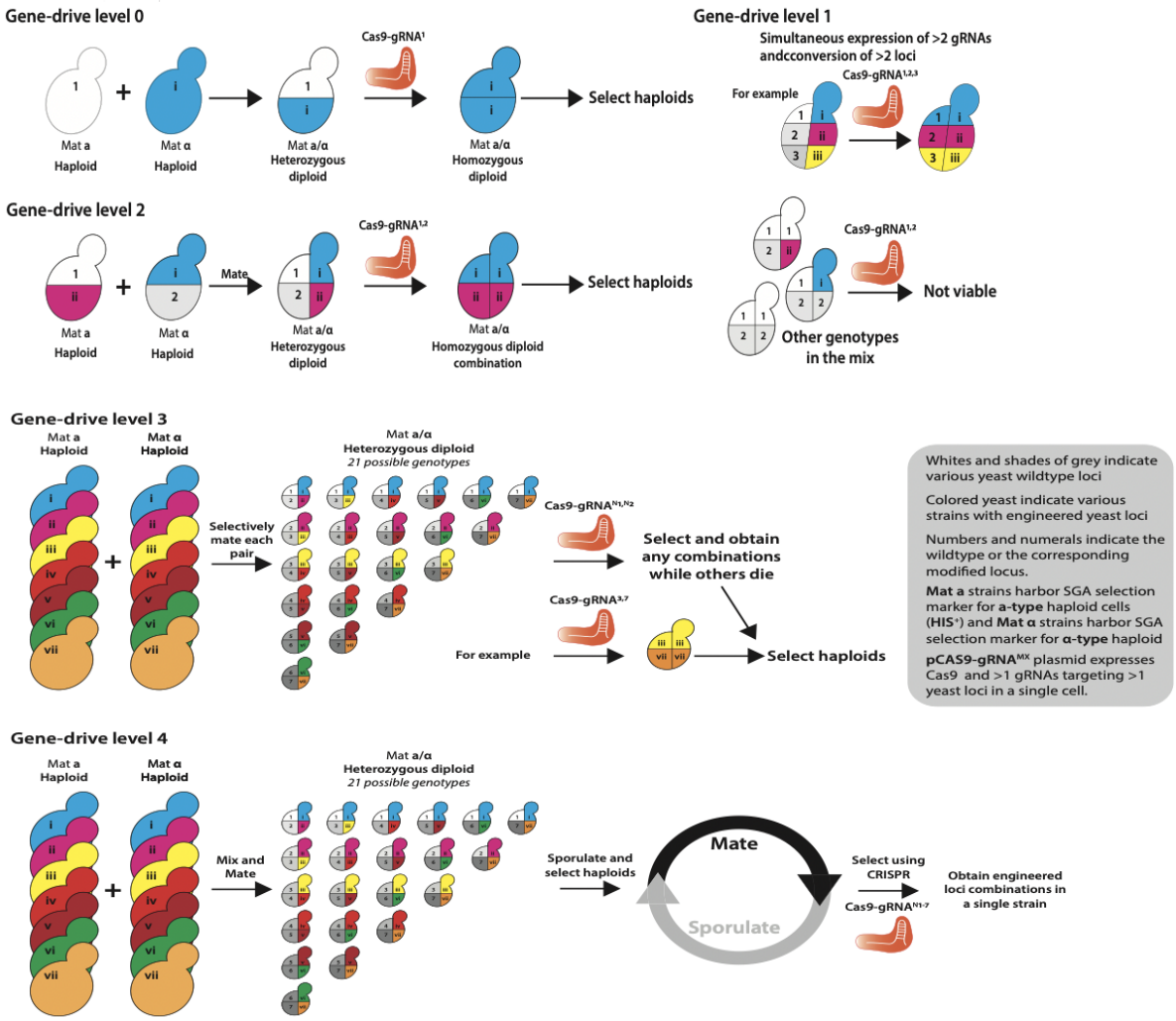


Figure 9. C-Gene Drive (CGD): CRISPR-Cas9 induced combinatorial gene editing methodology.

CGD level 0, shows simple conversion of any yeast locus to an engineered locus (colored) using CRISPR-Cas9-sgRNA selection. CGD level 1, involves similar conversion of >1 engineered loci. Both levels combine loci by mating engineered strains with wildtype strains. CGD level 2, shows combination of two independently generated engineered loci. In the case of CGD level 3, we propose to scale up level 2 to involve several engineered loci using Cas9-double-sgRNA plasmids as a selection to obtain doubly engineered homozygous genotypes. These levels combine genotypes by mating strains with two different engineered loci. And finally, in level 4, we follow level 3 strategy but perform several iterative cycles of mating and sporulation followed by Cas9-multi-sgRNA selection to obtain desired genotypes.

using a specific Cas9-multi-sgRNA expression vector that target multiple wildtype yeast loci and convert them to engineered loci in a single step. If successful, CGD **level 4** can help solve the bottlenecks when sequential humanization of a given combination reaches a dead end. Thus, the process will reveal a fitness-driven path to full humanization of yeast biological processes. The randomness of the mating and sporulating events can shed light on the important genetic or physical interactions that need to occur to allow even the "non-replaceable" modules to become replaceable.

4.2. Estimating the efficiency of CGD level 0 using yeast *Ade2* locus as a readout

In order for the CGD method to work optimally, each CGD level should work at high efficiency. To estimate the efficiency of the CGD **level 0,** a heterozygous diploid Magic Marker strain (*Ade2/Δade2::KanMX)* was used. This strain has three key properties that allow quick estimation of the CGD methodology. One, the strain provides a simple readout of the *Ade2* function as the disruption of loss-of-function of *Ade2* results in red colour colony phenotype [53]. Second, the strain carries an SGA marker (also referred to as the Magic Marker) that allows to easily select haploids spores from the mixture of unsporulated diploid yeast strains. Third, the strain carries two distinct alleles of *Ade2* yeast locus; a wildtype allele (*Ade2*) and a knockout allele (*Δade2::KanMX*). While the *Ade2* allele is prone to CRISPR-Cas9-sgRNA^{Ade2} mediated DSB, the *Δade2::KanMX* is resistant. Therefore, this *Ade2* locus should allow for an easy readout of conversion of after CGD **level 0**. We hypothesize that since an entire homologous chromosome serves as a repair template in a heterozygous diploid condition with one locus resistant to Cas9-sgRNA cut (i.e., an engineered locus) strains carrying a heterozygous engineered locus should become homozygous for that particular engineered locus at very high efficiency.

The transformation of the wildtype haploid or diploid cells (for *Ade2* locus) with pCRISPR-Cas9-sgRNA1^{Ade2} show a lethal phenotype with only few colonies (~0-20 CFU - referred to as **CFU$_O$**) surviving compared to the control plasmid (without sgRNA expression) transformation (~10^3 - 10^4 CFU - referred to as **CFU$_E$**) (**Figure 10A, I & II**). The estimated efficiency (in %) of the CRISPR experiment is measured by the formula (**CFU$_O$/CFU$_E$ x 100 = 0.74+/-SD; N=4**) (**Figure 10B, LEFT panel**). As previously explained, these few surviving colonies are a result of error-prone NHEJ and appear red suggesting the loss-of-function mutations at the *Ade2* locus. To test CGD **level 0**, we used a heterozygous knockout diploid *Ade2/ Δade2*::*KanMX* Magic Marker strain. We show, upon the transformation of pCRISPR-Cas9-sgRNA^{Ade2} plasmid targeting the

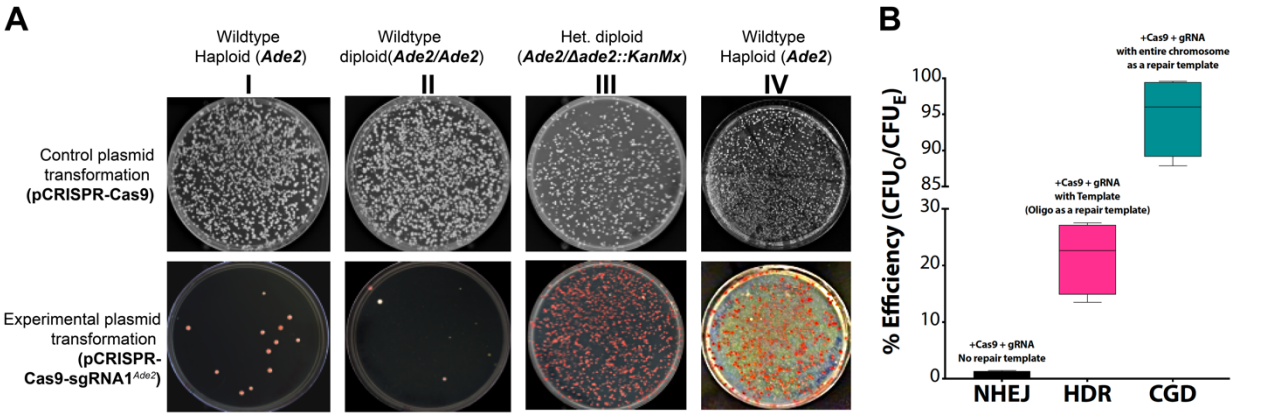


Figure 10. Estimating the efficiency of DSB repair at *Ade2* locus using CGD level 0

A. Haploid and diploid (homozygous and heterozygous for *Ade2*) yeast strains were transformed with control (**pCRISPR-Cas9**) and **pCRISPR-Cas9-sgRNA1^{Ade2}** plasmids. A DSB at the *Ade2* locus causes lethality in both the wildtype haploid and diploid cells with few survivors that repair the break via NHEJ showing a red colony phenotype (**I & II**). However, *Ade2/Δade2::KanMX* heterozygous yeast strain transformed with **pCRISPR-Cas9-sgRNA1^{Ade2}** shows nearly 100% viability of cells (similar to the empty vector transformation control) but all with the red colour phenotype (**III**). As an additional control, as shown in **IV**, transformation of **pCRISPR-Cas9-sgRNA1^{Ade2}** along with a repair template in the form of an oligo with 100bp sequence homology with the 5' and 3' UTRs of *Ade2* gene shows higher number of red colonies compared to the experiment without the repair template suggesting DSB repair via HDR. **B.** The data is represented in the form of a graph (N=4) as percent efficiency (**CFU$_O$/CFU$_E$**). In the absence of any repair template, yeast repair DSB via NHEJ (**LEFT panel**) with low error-prone repair efficiency. However, the number of colonies increases by **~5 fold** when a repair template is provided as an oligo (via HDR) (**MIDDLE panel**). On the other hand, the efficiency in the case of a **CGD level 0** is nearly 100% as every yeast cell that receives the CRISPR plasmid convert the locus to *Δade2::KanMX***.**

Ade2 locus, the conversion of *Ade2* to *Δade2*::*KanMX* at near 100% efficiency (**CFU_O/CFU_E x 100 = 95+/-SD; N=4**) (**Figure 10A, III & Figure 10B, RIGHT panel**). This experiment shows that nearly every yeast cell that receives the CRISPR plasmid is viable as the transformation efficiency is comparable to the empty vector transformation. As an additional control, the co-transformation of pCRISPR-Cas9-sgRNA1^{Ade2} and oligo as a repair template (with 100bp homology with the 5' and 3' UTRs of *Ade2* locus) in haploid wildtype yeast cells, show more survivors (**Figure 10A, IV & Figure 10B MIDDLE panel; CFU_O/CFU_E x 100 = 21.6+/-SD; N=4**) i.e., ~30 fold higher efficiency than an error-prone NHEJ pathway. However, this method is still far less-efficient than CGD **level 0 (Figure 10B; 95 vs 21.6% efficiency)**.

However, the red colonies observed in CGD **level 0** could be due the mutation in the *Ade2* locus (NHEJ pathway) or due to conversion to Δ*ade2*::*KanMX* locus (HDR pathway) (**Figure 11A**). To verify that the all the red colonies have converted an *Ade2* to Δ*ade2::KanMX* locus, the heterozygous diploid (*Ade2*/*Δade2*::*KanMX*) strain transformed with a control (pCRISPR-Cas9) and pCRISPR-Cas9-sgRNA^{Ade2} plasmid were sporulated for tetrad dissection. As observed in **Figure 11B**, spores harbouring a control plasmid show 2:2 red:white phenotype and 2:0 G418 resistant: G418 sensitive phenotype. This was expected as control plasmids do not express a sgRNA therefore do not create DSB at the *Ade2* locus. However, in the case of pCRISPR-Cas9-sgRNA^{Ade2} transformation, as previously hypothesised, the repair of the DSB should occur primarily via HDR as the yeast cell repairs DSB using the homologous chromosome as a repair template (**Figure 11A**). To investigate that, I performed tetrad dissection and replica plating for antibiotic resistance for the pCRISPR-Cas9-sgRNA^{Ade2} plasmid transformed condition as well. Tetrad dissection shows all 4 spores with 4:0 red:white colonies. Replica plating the colonies on YPD+G418 plate confirmed the results indicating 4:0 G418 resistant:G418 sensitive colonies

suggesting the conversion to *Δade2::KanMX* rather than the mutation of *Ade2* locus (**Figure 11B**).

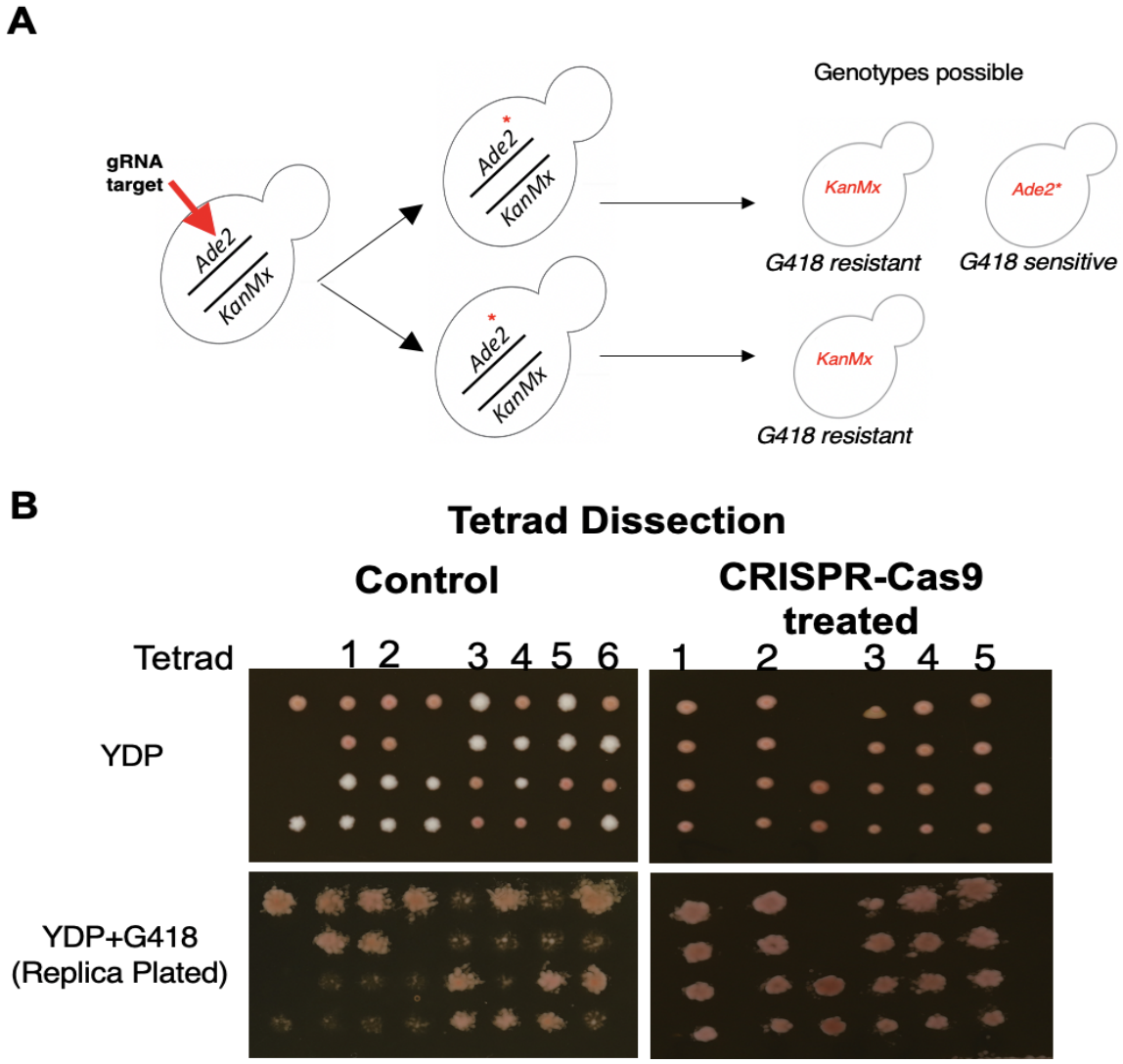


Figure 11. CGD mediated conversion from *Ade2* to *Δade2::KanMX* locus is near 100% efficient.

A. Schematic of possible genotypes when a heterozygous diploid *Ade2/Δade2::KanMX* strain is subjected to a Cas9-sgRNA mediated DSB followed by sporulation. The DSB repair via NHEJ could simply mutate the locus (*Ade2**) or repair via HDR to convert the locus to *Δade2::KanMX*. Either scenarios lead to a red colony phenotype. However, if the strain is sporulated, the prior scenario will show only 50% G418 resistant colonies compared to the later outcome that will exhibit 100% G418 resistance. **B.** Tetrad dissection of pooled cells from heterozygous diploid strains transformed with either a control plasmid or with pCRISPR-Cas9-sgRNA^{Ade2}. Control plasmid experiment shows that 2:2 red:white colony phenotype. However, pCRISPR-Cas9-sgRNA^{Ade2} transformed cells show 4:0 red:white colony phenotype. Furthermore, all 4 spores in the case of CRISPRed strain show 100% G418 resistance.

4.3. Testing the efficiency of CGD level 0 at essential yeast loci

While we showed the conversion of a non-essential yeast *Ade2* to engineered locus, we wished to similarly verify the efficiency of CGD **level 0** for any essential yeast locus as well. Therefore, as a representative example, we used Sc*Alpha4* proteasome gene, a subunit of the proteasome core. Using the CRISPR-Cas9 based method of humanization described in **Chapter 2 and 3**, we had previously successfully humanized the yeast Sc*Alpha4* (with human Hs*PSMA6*) (data from Abdullah. M, not shown). We first show that the humanized strain *ScAlpha4*::Hs*PSMA*6 was resistant to the DSB induced by pCRISPR-Cas9-sgRNA^{Alpha4} further verifying the engineered locus (**Figure 12A, Second from left**). The number of CFUs for control (pCRISPR-Cas9) and experimental pCRISPR-Cas9-sgRNA^{Alpha4} plasmid transformed were similar (**Figure 12A, first and second plates from left**). Next, this haploid Sc*Alpha4*::Hs*PSMA6* Mat-A strain was mated with a wildtype Mat-α strain to create a heterozygous diploid harboring both a yeast and human gene. We transformed both the heterozygous diploid *ScAlpha4*/ Sc*Alpha4*::Hs*PSMA6* strain and the diploid wildtype strain (*ScAlpha4*/Sc*Alpha4*) with pCRISPR-Cas9-sgRNA^{Alpha4} *[Refer to Materials and Methods, section 7.2*]. As seen in **Figure 12A** (**third plate from left**), the wildtype diploid showed lethality (very few survivors) indicating that a DSB on both copies of the yeast gene was not repaired efficiently. However, in the case of the humanized heterozygous diploid strain, the number of colonies obtained after transformation with pCRISPR-Cas9- sgRNA^{Alpha4} plasmid showed comparable CFUs to the empty vector control suggesting viability and efficient repair of the DSB via HDR using the homologous chromosome as a repair template (**Figure 12A, fourth & fifth plate from left**). The data shown as % efficiency (CFU_O/CFU_E) in **Figure 12B**, mimics the data obtained in the case of *Ade2* CGD (**Figure 10B**), suggesting the CGD **level 0** is equally efficient at essential yeast loci as well. The DSB repair via

HDR using PCR as a repair template (with 80-100bp homology to 5' and 3' UTRs) is far less efficient than the CGD (**Figure 12B, MIDDLE panel)**.

To further verify the conversion of loci, yeast colonies from the CRISPRed heterozygous strain were sporulated, tetrad dissected and genotyped to confirm the locus conversion from

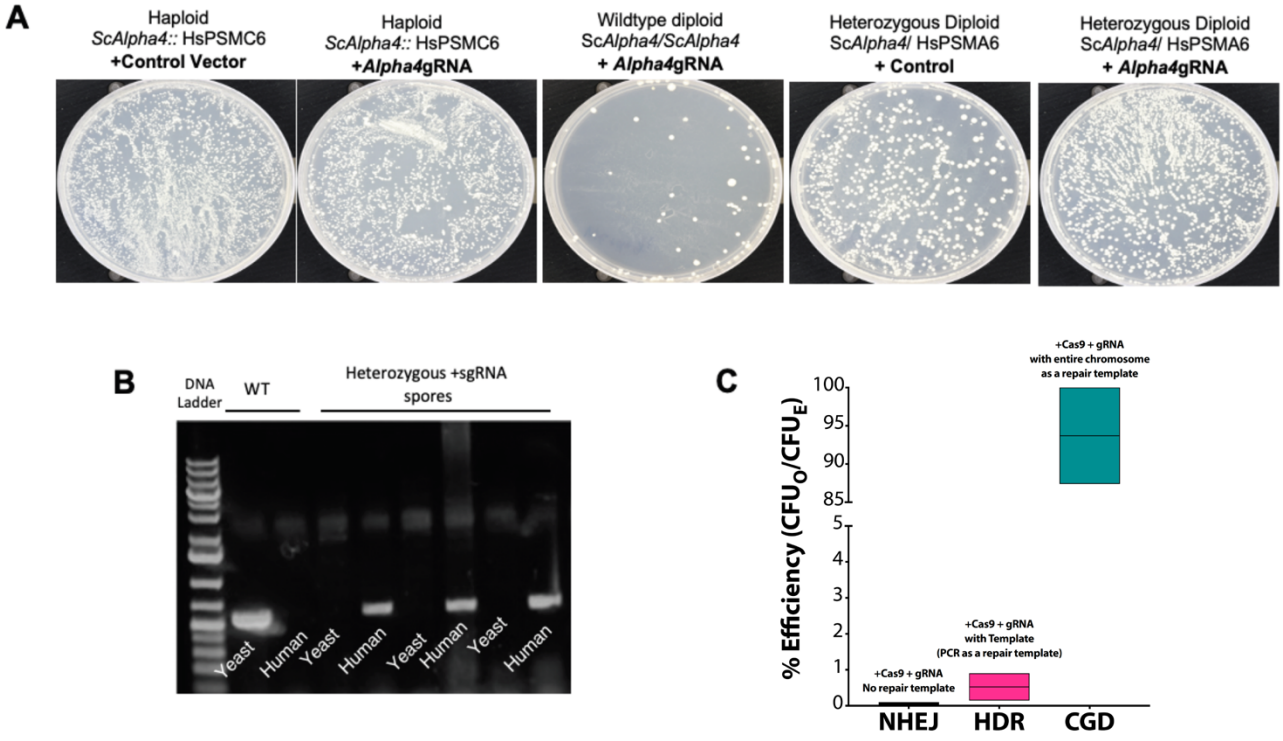


Figure 12. CGD level 0 conversion of essential yeast locus is as near 100% efficient.

A. Various yeast strain transformations show the efficiency of the pCRISPR-Cas9-sgRNA^{Alpha4} mediated CGD **level 0** targeting the essential yeast gene Sc*Alpha4*. A haploid wildtype strain transformed with a control plasmid serves as a positive control for transformation efficiency (first panel from left, **CFUO**). A haploid yeast strain harbouring humanized locus (Sc*Alpha4::HsPSMA6)* shows resistance to DSB when transformed with pCRISPR-Cas9-sgRNA^{Alpha4} plasmid (second from left). As is indicated by increased number of colonies similar to the control plasmid transformation. The wildtype diploid yeast cell shows lethality upon transformation with pCRISPR-Cas9-sgRNA^{Alpha4} plasmid (third from left). However, heterozygous *ScAlpha4 / ScAlpha4::*Hs*PSMA6* diploid when transformed with pCRISPR-Cas9-sgRNA^{Alpha4} (**CFUE**, fifth from left) shows comparable CFUs to the empty vector transformation (fourth from left). **B.** Colony PCR confirmed the conversion of the heterozygous to a homozygous locus. Wildtype control shows the amplification of yeast Sc*Alpha4* locus. However, all randomly picked colonies from a CGD **level 0** (fifth from left) strain showed the conversion of the yeast to human locus. **C.** The data is represented in the form of a graph (N=3) as percent efficiency (**CFUO/CFUE**). In the absence of any repair template, yeast repair DSB via NHEJ (**LEFT panel**). However, the number of surviving colonies increases when a repair template is provided as an PCR (via HDR) (**MIDDLE panel**). Comparatively, the efficiency in the case of a **CGD level 0** is nearly 100% as every yeast cell that receives the CRISPR plasmid converts the locus to *ScAlpha4::*Hs*PSMA6*. **(Data from Adbullah,M.)**

yeast/human to human/human. All three randomly picked colonies showed the conversion of the yeast to human gene locus (**Figure 12B**).

4.4. Using SGA selection to efficiently select genotypes with engineered loci for CGD

The strains used in this part of **Chapter 4** were obtained from Dr. Charlie Boone at the University of Toronto [33]. These strains allow selection of each haploid mating type from a mixture of haploid and diploid cells. A diploid strain with the *can1delta::STE2pr-Sp_his5 / lyp1delta::STE3pr-LEU2; his3delta1 leu2delta0 ura3delta0* genotype was sporulated to obtain haploid strains of both mating types with SGA (Synthetic Genetic Array) markers for easy haploid selection. The SGA markers are inserted at *Can1* and *Lyp1* loci, thereby, making the strains resistant to canavanine (CAN) and thyalysine (THY) respectively. The CAN resistance comes from the *can1* locus deletion where the *can1* gene encodes for an arginine receptor. Since canavanine is structurally similar to arginine, it would enter through this receptor and kill the cell, but with this deletion the cell can survive the presence of CAN in the medium (**Figure 13A**) [33]. Similarly, the *lyp1* deletion allows the thialysine resistance as the *lyp1* gene encodes a lysine receptor. This diploid strain is sensitive to CAN and THY as one functional copy of *Can1* and *Lyp1* genes are still intact **(Figure 13A).** The inserted cassettes carry auxotrophic selection markers driven by haploid-specific promoters. The *can1delta::STE2pr-Sp_his5* cassette allows the selection of Mat-A haploid cells (HIS+, CAN+) whereas the *lyp1delta::STE3pr-LEU2* cassette allows the selection of Mat- α haploid cells (LEU+, THY+). The *STE2p* promoter, a Mat-A specific promoter, drives the expression of *His5* gene, thus allowing the selection of these haploid cells on SD-HIS+CAN medium. The *His5* gene used in this strain is a *Schizosaccharomyces pombe* specific gene to prevent recombination with the *Saccharomyces*

cerevisiae His gene in the strain [56]. Similarly, selection on SD-LEU+THY allows the selection for Mat- α haploid cells (*STE3pr is a MAT- α* specific promoter*)*(**Figure 13A**). These SGA strains are also compatible for mating with other non-SGA strains (**Figure 13B & D**). Therefore, CGD **level 0** strategy allows the conversion of wildtype SGA strains to engineered strains by mating, sporulating and targeting specific loci with CRISPR-Cas9-sgRNA complex. Humanized BY4741 hapliod strains (Mat-A) were converted to SGA strains using plasmid-based mating [*Refer to Materials and Methods, section 7.4]* and mated heterozygous diploids were confirmed by colony PCR (**Figure13B & C**). The strains were sporulated and tetrad dissection was performed followed by replica plating onto SGA media (Mat- α -Leu+THY) to show only 1 of 4 haploid spores grow (**Figure 13D**). The data for humanized SGA strains is not shown.

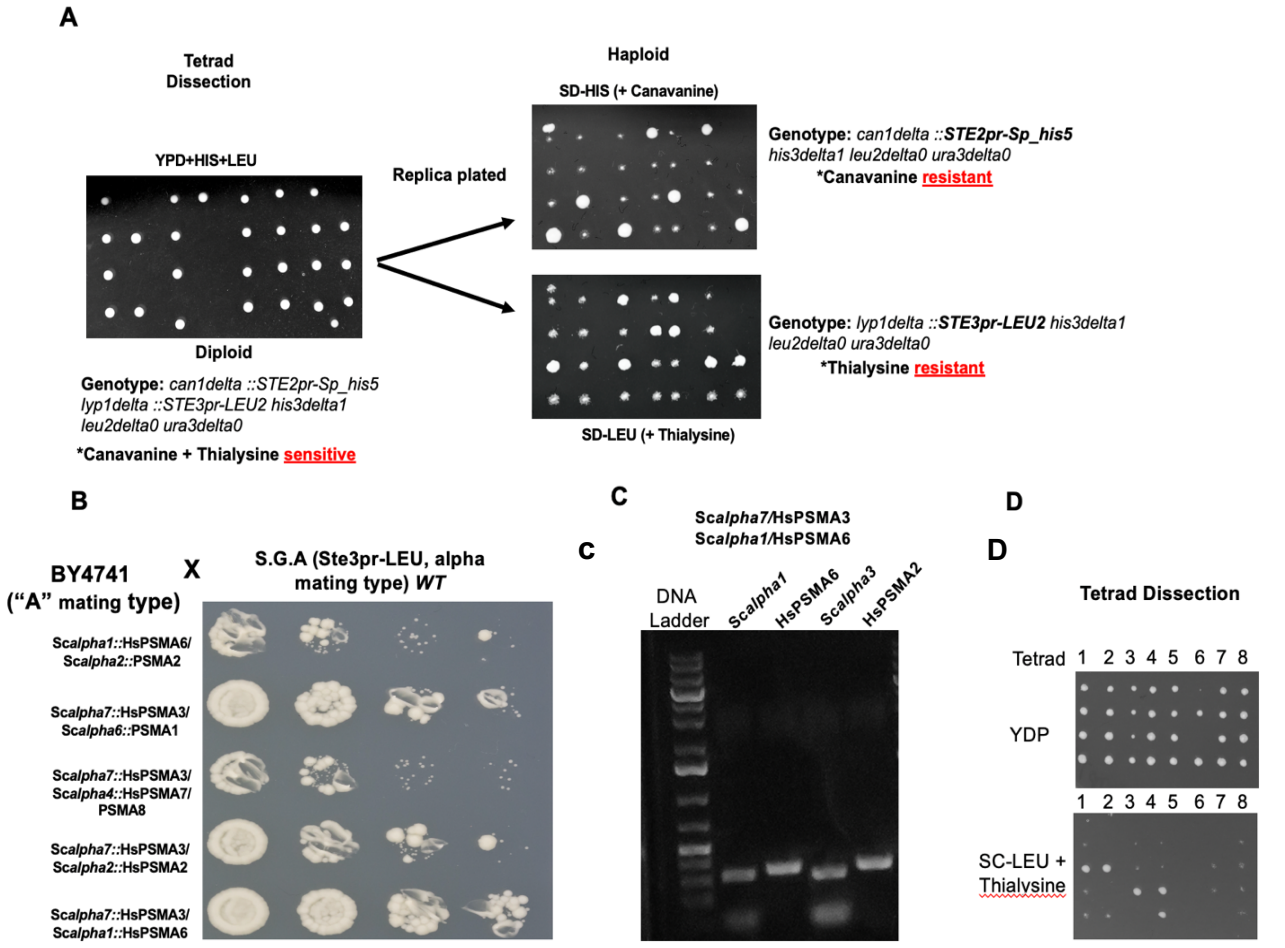
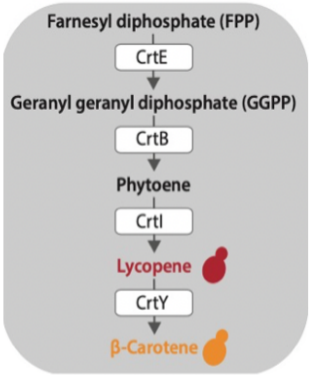


Figure 13. Construction of engineered humanized SGA selection strains that allow for easy haploid selection
A.A diploid yeast strain harboring SGA markers shows canavanine (CAN) and thialysine (THY) sensitivity. After sporulation and tetrad dissection, haploid spores were replica plated on SD-His+CAN to select for mating type A strains and on SD-Leu+THY to select for the mating type α strains. **B.** Plasmids that provide different selections were transformed in each mating type strain. The doubly humanized strains (Mat-A) were mated with SGA strains (Mat- α). This plasmid-mediated selection shows the efficiency of SGA strains (Mat- α) to mate with humanized BY4741 of opposite mating type (Mat-A). Growth on double selection indicates mating occurred. **C**. A representative colony PCR of a double humanized strain (*ScAlpha1*/Sc*Alpha1*::Hs*PSMA6*; *ScAlpha7*/Sc*Alpha7*::Hs*PSMA3*) confirms the heterozygous nature of the humanized and wildtype yeast loci. **D.** Tetrad dissection of a mated SGA X BY4741 strains show the selection via SGA marker as only 1 of haploid spores is growing on selection media. **(Data from Adbullah,M.)**

4.5. Building heterologous pathways in yeast using CGD

As a proof of principle for CGD **level 4**, we are in the process of combining the entire heterologous carotenoid pathway (4 gene system: *CrtE-->CrtI-->CrtB-->CrtY/YB*) in yeast. Carotenoid pathway provides an easy readout of correct assembly as the intermediates show coloured colony phenotype. When the first three genes of the pathway are expressed, the strain produces lycopene which is red in colour whereas the entire pathway assembly produces β-carotene which is orange in colour [50–52] (**Figure 14A**). Carotenoid pathway genes were obtained from the *Pantoea ananatis* (bacteria) (*CrtE* and *CrtI*) and *Xanthophyllomyces dendrorhous* (yeast) (*CrtB* and *CrtY*). The four genes that comprise carotenoid pathway (**Figure 14A & B**) were individually engineered into a non-coding safe harbor sites in a haploid SGA (Mat-A) yeast strain that is easily selectable with an auxotrophic marker (HIS) (**Figure 14C**). These integration of these four genes were confirmed via colony PCR using engineered loci specific primers (**Figure 14D**).

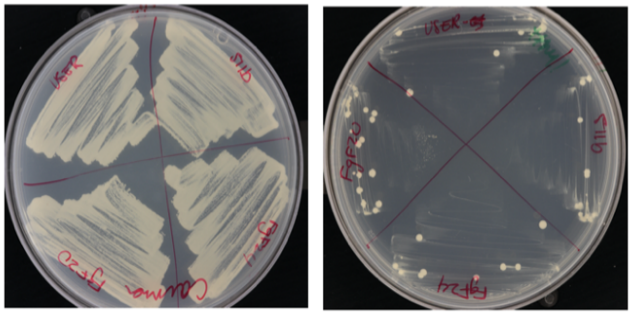
A



B

Yeast Locus	Carotenoid Gene
511b	CrtE
USERX-1	CrtB
FgF20	CrtI
FgF24	CrtYB/Y

C



D

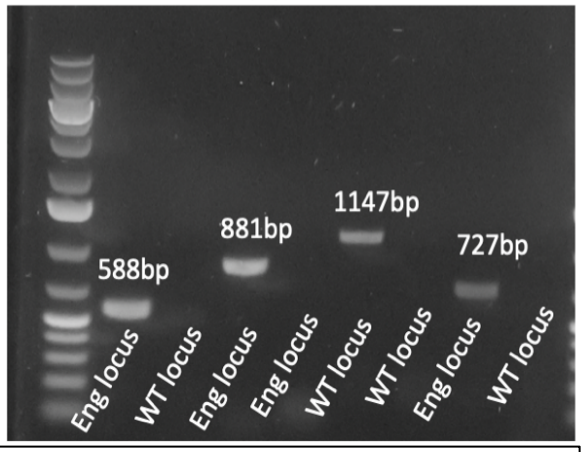
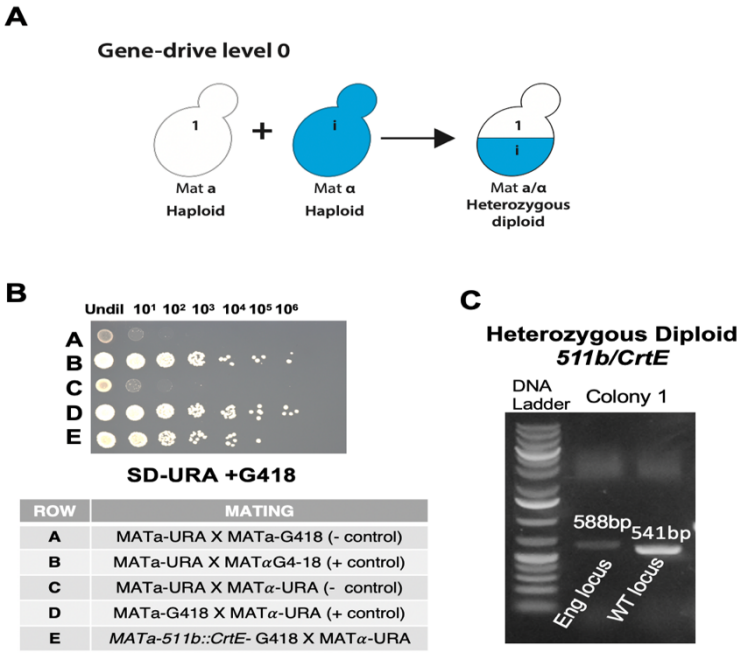


Figure 14. Carotenoid pathway gene integrations in SGA strains.

A. Schematics of the complete carotenoid biosynthesis pathway. When the first three genes of the pathway are assembled, the strain makes lycopene (red pigment) and when the full pathway is assembled, it makes carotene (orange pigment). **B.** Carotenoid genes were integrated into non-coding safe-harbour yeast loci. **C.** The carotenoid genes were integrated into the SGA strain (Mat-A) that allows growth on SD-HIS+CAN and not on SD-LEU+THY. **D.** Colony PCR results confirm the integration of carotenoid genes in SGA strains. **(Data in collaboration with Drs. Pyne M & Martin V.)**

Next, the engineered strains were mated with a WT tester strain (Mat- α, SGA-LEU+THY) to create heterozygous diploids for each of the carotenoid genes using a plasmid-based selection as previously described (**Figure 15B**). One strain harbouring a plasmid with a URA selection of one mating type was mixed with another strain of the opposite mating type with a G418 vector and plated on double selection (SD-URA +G418) to ensure that only mated cells will survive. As a representative example, I show the results for the first gene of the carotenoid pathway (*CrtE)* integrated at the *511b* safe harbour yeast locus. After mating the strain shows the presence of both engineered and the wildtype locus specific PCR (**Figure 15C**). Colonies were picked and confirmed for heterozygous nature of the locus in diploids where both the yeast and engineered locus specific primers will bind and provide a positive PCR fragment. Heterozygous diploids for the rest of the three genes of the pathway were generated in a similar fashion (**FigureS1**).



ROW	MATING
A	MATa-URA X MATa-G418 (- control)
B	MATa-URA X MATαG4-18 (+ control)
C	MATa-URA X MATα-URA (- control)
D	MATa-G418 X MATα-URA (+ control)
E	*MATa-511b::CrtE*- G418 X MATα-URA

Figure 15. Generating heterozygous diploids with SGA markers to build the Carotenoid pathway in yeast using CGD.
A. Schematic of creating heterozygous diploid for wildtype and engineered loci to be used for gene drive level 0. **B.** Singly engineered carotenoid strain (*511b::CrtE*) was mated with wildtype SGA strain to generate heterozygous diploids (plasmid-based mating). Table below indicates the control and experimental mating reactions in each row. **C.** A colony PCR from mating plate confirms the heterozygous nature of the locus**.**

Next, using CGD **level 0**, homozygous diploids were generated by targeting each individual WT locus with a CRISPR-Cas9-sgRNA system to create homozygous diploids (**Figure 16A**). For example, in the case of pCRISPR-Cas9-sgRNA^{511b} transformed in the *511b/511b::CrtE* strain, CRISPR targets the intact *511b* locus To ensure the pCRISPR-Cas9-sgRNA^{511b} mediated DSB specificity, a haploid WT cell transformed with a pCRISPR-Cas9-sgRNA^{511b} resulted in lethality compared to the control transformation (**Figure 16B, first and second from left**). The same plasmid when transformed in heterozygous diploid for *511b/511b::CrtE* showed no lethality (CFU similar to the control transformed experiment) and resulted in a homozygous diploid strain (**Figure 16B, third and fourth from left**). All four homozygous strains were PCR confirmed (**Figure S2**), mixed, mated and sporulated. The CGD **level 4** as illustrated in **Figure 17 is in progress**. The cycle of mating and sporulating can allow cells to combine loci while the specific CRISPR-Cas9-multi-sgRNA complex will allow the selection for a specific combination of engineered genotypes.

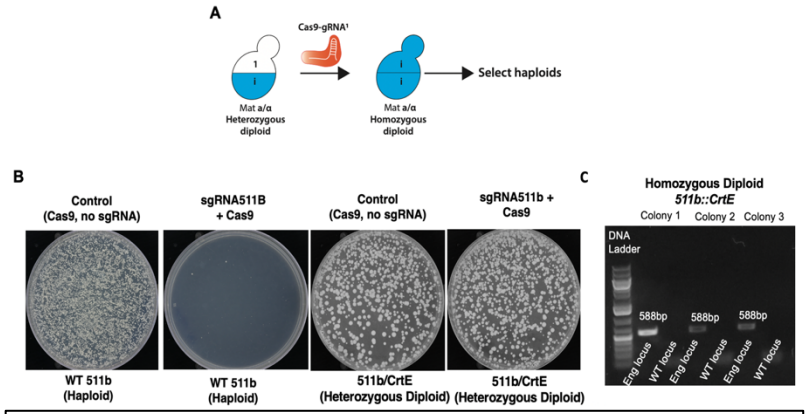


Figure16. CGD level 0 conversion of yeast locus to engineered carotenoid gene locus.

A. CGD level 0 schematic showing the use of CRISPR-Cas9 to create a homozygous diploid strain for carotenoid genes after mating with opposite mating-type wildtype SGA strains. **B**. A representative example shows upon transfortion of a pCRISPR-Cas9-sgRNA^{511b} plasmid, the surviving cells convert a *511b* to an engineered locus. Heterozygous diploid (*511b/511b::CrtE*) transformed with pCRISPR-Cas9 -sgRNA^{511b} plasmid shows no lethality as the number of CFUs is similar to the empty vector transformed control. **C.** Randomly picked colonies from the *511b/ 511b/CrtE* heterozygous diploid strain transformed with pCRISPR-Cas9-sgRNA^{511b} plasmid show homozygous engineered locus as only engineered locus specific PCR bands are observed.

Depending on the number of engineered loci, several cycles of mating and sporulation may be needed to assemble the entire engineered pathway (for example, for a 4-gene carotenoid pathway starting from haploid strains, a minimum of 2 mating and sporulation cycle should yield a fully assembled pathway in a single strain) (**Figure 17**). Based on the colony color phenotype, we can easily screen for a fully assembled engineered strain. As expected several other genotypes will be present in the mix. However, using CGD and targeting >1 yeast locus can allow for the selection of specific genotypes. For example, only a strain with the first three carotenoid genes

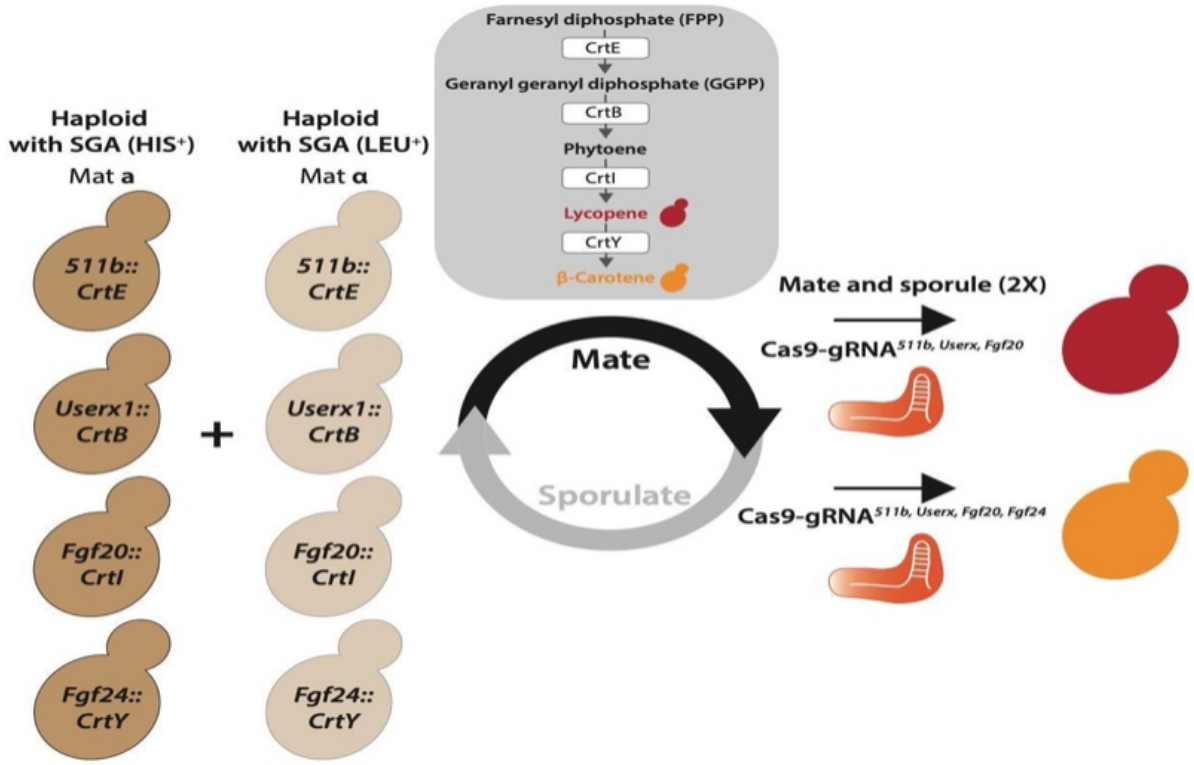


Figure 17. CGD level 4 methodology for engineering the carotenoid pathway in yeast..

Homozygous haploid SGA strains for each carotenoid gene are mated as a pool. Next, mated cells are sporulated and the process repeated for at least 2 rounds. This pool of haploid and diploid yeast cells with various genotypes are then transformed with specific CRISPR-Cas9 plasmids targeting different yeast loci. The CRISPR-Cas9-sgRNA plasmids will allow the selection of specific engineered genotypes respectively while eliminating other genotypes. For example, haploid or diploid cells harboring the first 3 or all 4 carotenoid genes can be selected with 3 or 4 gRNA expressing plasmids respectively. The colored colony outcome will further serve as an easy readout of the correct combination of engineered genotypes.

either in homozygous or heterozygous state will survive a pCRISPR-Cas9-sgRNA$^{511b;UserX\text{-}1;Fgf20}$ cut that targets the three WT yeast loci respectively. The CRISPR-Cas9-sgRNA system thus serves as an efficient selection to eliminate any other unwanted genotypes (**Figure 17**).

5. Chapter 5 - Conclusion

My Master's thesis project focused on the development of a next-gen tool to build heterologous pathways, including humanized systems, in a simplified eukaryotic cell. Yeast and humans share >4000 genes and many shared yeast genes are swappable by their human equivalent's (one-gene-at-a-time) [31]. These systematic studies allow us to explore the ultimate goal of full humanization of genetic systems in yeast. Here, we provide a roadmap of a method to integrate human systems at their native yeast loci and convert/combine humanized loci using CRISPR-Cas9. We have successfully individually humanized genes involved in metabolic pathways and multi-gene complexes using CRISPR-Cas9 [16,26]. Though singly humanized yeast strains are valuable tools to study, most genes do not work in isolation, but rather as systems composed of multiple genes. While several methods allow engineering of biological systems in yeast, we propose a novel, scalable method using CRISPR-Cas9 as a selection for combination of engineered loci.

In **Chapter 2**, I constructed several CRISPR-Cas9 reagents targeting yeast loci and tested for their efficacy by introducing these reagents into haploid yeast cells. Efficiency of CRISPR-Cas9 reagents was evaluated based on a lethal phenotype as haploid yeast cells. The lethal phenotype is likely due to the majority of yeast cells that repair the locus via non-error prone NHEJ pathway [20,21,57]. The regeneration of CRISPR site after efficient NHEJ results in continuous targeting of the locus by Cas9-sgRNA complex. The cells that do mutate the locus due to an error prone NHEJ pathway are far fewer (due to less efficient error-prone pathway), thus, resulting in only few survivors depending on the essentiality of the locus.

Once these CRISPR reagents were created, tested and sequence verified, I used them to singly replace yeast genes with their human counterparts for a subset of genes (*Rpt1-6)* that make up the proteasome base (**Chapter 3**). Three of six proteasome base genes (*Rpt3, 5,6)* were successfully humanized into the yeast genome. These strains were characterized through PCR-based genotyping, sequence verification and examined for fitness defects [*Refer to Materials and Methods, sections 7.3, 7.5, 7.6].* In **Chapter 4,** we developed a method that uses CRISPR-Cas9 reagents, mating and sporulate cycles as a means to combine and convert loci to build entire engineered systems. As a proof of principle, I showed the efficiency of conversion of a non-essential *Ade2* locus followed by the combination and conversion of essential yeast loci (core of the proteasome). To show the general application of the C-Gene Drive (CGD) method, I show the preliminary data for the assembly of heterologous carotenoid pathway (*CrtE, CrtB, CrtI, CrtYB*) in yeast. As of now, individual strains containing each of the four gene have entered the mating and sporulating cycles (CGD **level 4**) and data collection is in progress.

This method works because each CRISPR-Cas9 reagent induces a lethal DSB in a wildtype yeast locus but not in engineered loci while the entire homologous chromosome serves as an efficient and stable repair template for HDR. Using our collection of CRISPR vectors, we were able to integrate several human genes in the yeast genome and convert the loci into SGA strains for efficient haploid specific selection. Next, we plan to use CGD to fully assemble the entirely human biological processes in yeast.

Several methods allow engineering of complex biological processes in yeast [33,34]; methods such as using YACs to clone entire pathway, plasmid shuttling and sequential assembly of modules [35–37]. Though these methods are powerful, they are not compatible with engineering humanized systems in yeast. Our method to combine loci requires no deletions for

selection but rather uses CRISPR-Cas9 to select for specific combination of human loci driven by fitness. Only fitter humanized strains will survive or be represented at higher frequency in a pool, thus, allowing the characterization of the mechanisms of engineered pathways by testing genetic or physical interactions in hybrid human/yeast strains. Going forward, we hope to build fully humanized systems directed by the fitness of the intermediate hybrid strains potentially shedding light on the incompatibilities of human/yeast hybrid systems. These strains will serve as biological reagents to test the contribution of genetic variations associated with Mendelian and complex human diseases.

While we intend to use the CGD methodology to humanize yeast biological processes in their entirety, the method is broadly applicable to engineer any heterologous pathway in yeast (we show it for the carotenoid pathway). Furthermore, since the method allows selection-less combination of engineered yeast loci, it can also allow combining multiple gene deletions to test higher-order complex genetic interactions in yeast. The only limiting factor being the co-expression and assembly of many Cas9-sgRNA complexes in a single yeast strain.

6. Chapter6- Materials and Methods

6.1. Strains & media

All yeast strains were grown in YPD, or in synthetic dropout media with or without selection or in rich GNA media. The following media components from Sigma Aldrich were used: YPD powder, yeast synthetic drop-out medium supplements without uracil, dextrose, agar, L-glutamic acid monosodium salt hydrate, yeast extract, nutrient broth, thialysine (60µg/ml), cannavinine (60µg/ml) and G418/Geneticin (200µg/ml) from Gibco.

The haploid yeast strains used are BY4741 with the genotype *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* and BY4742 with the same genotype with a MATalpha mating type. Diploid strains from SGA collection from the Boone lab were used the genotype MATa*/a ura3D0 leu2D0 his3D1 lys2D0/LYS+ met15D0/MET15+ can1delta::LEU2+-MFA1pr-HIS3/CAN1+ xxx::kanMX/XXX+* where *XXX+* is the gene of interest and in this case is *Ade2* and *Dgf10* genes. The other SGA diploid strain used was Y8205 with the genotype *MAT* alpha *can1delta:: STE2pr-Sp_his5 lyp1delta ::STE3pr-LEU2 his3delta1 leu2delta0 ura3delta0* [33].

6.2. Plasmid preparation and yeast transformation

The CRISPR-Cas9-sgRNA plasmids were transformed into competent *E. coli* cells. Plasmids were then isolated from *E. coli* using the Monarch Plasmid MiniPrep kit from New England BioLabs. The CRISPR-Cas9-sgRNA plasmids were then transformed into BY4147 and By4742 haploid cells that were made competent using the Frozen-EZ Yeast Transformation II Kit from Zymo Research. Haploid yeast cells were transformed with 1µg of CRISPR-Cas9-sgRNA plasmid and 2-5µg of human template (linear PCR fragment). Human templates were amplified from either a Dharmacon or HumeORF collection using hybrid primers [54,55]. Hybrid primers

are made up of 20bp homology to the human gene of interest and 80bp of yeast homology of promoter and terminator regions. As a control, a CEN6GPF-Ura/G418 plasmid was used where a Cas9 cassette is present but a sgRNA cassette is absent. Transformed cells are plated on appropriate selection media and incubated at 30ºC for 3-5 days.

6.3. Confirmation of singly humanized strains

Engineered yeast colonies were treated with 15µl 50mg/ml zymolase (US biologicals) and 5µl of water and heated [16]. A forward primer homologous to regions just outside the flanking left- homologies of the gene of interest locus and an internal reverse primer binding to the yeast or the human gene were used to amplify using PCR to verify the changes at that region. The following protocol was used for the PCR reaction. The lysate was exposed to an initial denaturation step of 2 min at 95°C, followed by 30 cycles of denaturation at 95°C for 30 seconds, 30 seconds of primer annealing at 55°C and 2 minutes of extension at 68°C. The cycle was followed by a final step of 10 minutes at 68°C to ensure all elongation reactions were complete. We used *AccuPrime Pfx* DNA polymerase (Invitrogen) to carry out the PCR.

6.4. Mating, sporulation and haploid selection cycle

Mating procedures using a plasmid-based selection was used where parental strains used were opposite mating types (MATa vs. MATalpha). Each parental strain was transformed independently with two different plasmids with two different selectable markers (URA and G418). A colony from each transformed strain was mixed in a rich medium such as YDP and incubated at 30ºC for 4-24h with shaking. Mix was then diluted and plated onto double selection media (SD-URA + G418). Plasmids were cured by repeated streaking onto rich YDP medium and counter selected on 5-FOA for strains harbouring a URA plasmid. Strains were then

inoculated in 1-5ml of GNA overnight and then washed for 500µl to 1ml of sporulation media (100ul of 10%Potassium acetate (Sigma P1190), 10µl of 0.5% Zinc acetate (Sigma Z0625) and 890ul of autoclaved MiliQ water). After washing, cells were suspended in 500ul of sporulation media and 200ul of the mix was diluted into 1ml of sporulation media. Cells were exposed to rotation and shaking (120-140rpm) at room temperature for 3-7 days. Sporulation mix was analyzed for spores under a microscope and either prepared for tetrad dissection or haploid selection. For tetrad dissection, spores were spun down at 5000rpm for 5 minutes, resuspended in 200µl of 20mg/ml zymolase and incubated at 37ºC for 25-30 minutes. The mix was then put in -20ºC to stop the reaction. Once frozen, cells were thawed on ice and 20ul of mix was plated into YDP for tetrad dissection. To select for haploids, spore mix was incubated in Sc-HIS+ CAN and Sc-LEU+ THY overnight independently. Equal amounts of both overnight cultures were mixed in YDP to restart the mating/sporulation cycle.

6.5. Growth curve analysis

All strains were grown in YDP at 30ºC for 48 hours in 96 well Corning plates. One microliter of overnight culture was diluted in 149µl of YDP. Blanks consisted of water and YDP and wildtype strains were either BY4741 or BY4742. Tecan Sunrises machines were used for these assays. Data was analyzed using the Graph Prism software. Error bars were derived from standard deviations between technical replicates.

6.6 Single CRISPR and CGD efficiency analysis

Percent efficiency of CRISPR experiment was calculated by dividing the number of viable cells observed (CFU_O - O stands for observed) when transformed with a specific CRISPR-

Cas9+sgRNA plasmid (+/-repair template) by the number of cells obtained after a control plasmid (without sgRNA expression) transformation (CFU_E - E stands for expected). Percentage efficiency was calculated by the following formula: CFU_O / CFU_E x 100.

7. Chapter 7- Bibliography

1 Kumar, A. and Snyder, M. Emerging technologies in yeast genomics. , *Nature Reviews Genetics*, 2. (2001) , 302–312

2 Cooduvalli S Shashikant, Janet L. Carr, Jaya Bhargava, Kevin L Bentley, Frank H Ruddle (1998) Recombinogenic targeting: a new approach to genomic analysis—a review. *Gene* 223, 9–20

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

4 Hinnen, A. *et al.* (1978) Transformation of yeast. *Proc. Natl. Acad. Sci. U. S. A.* 75, 1929–1933

5 Scherer, S. and Davis, R.W. (1979) Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 76, 4951–4955

6 Orr-Weaver, T.L. *et al.* (1981) Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. U. S. A.* 78, 6354–6358

7 Ding, Y. *et al.* (2019) Increasing the homologous recombination efficiency of eukaryotic microorganisms for enhanced genome engineering. *Appl. Microbiol. Biotechnol.* 103, 4313–4324

8 Storici, F. *et al.* (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14994–14999

9 Rudin, N. and Haber, J.E. Efficient repair of HO-induced chromosomal breaks in Saccharomyces cerevisiae by recombination between flanking homologous sequences. , *Molecular and Cellular Biology*, 8. (1988) , 3918–3928

10 Plessis, A. *et al.* (1992) Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* 130, 451–460

11 Storici, F. and Resnick, M.A. (2006) The Delitto Perfetto Approach to In Vivo Site-Directed Mutagenesis and Chromosome Rearrangements with Synthetic Oligonucleotides in Yeast. In *Methods in Enzymology* 409pp. 329–345, Academic Press

12 Yellman, C.M. (2020) Precise Replacement of Saccharomyces cerevisiae Proteasome Genes with Human Orthologs by an Integrative Targeting Method. *G3* 10, 3189–3200

13 Mojica, F.J. *et al.* (1993) Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* 9, 613–621

14 Sternberg, S.H. and Doudna, J.A. (2015) Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol. Cell* 58, 568–574

15 Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. , *Science*, 337. (2012) , 816–821

16 Akhmetov, A. *et al.* (2018) Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9. *Bio Protoc* 8,

17 Kachroo, A.H. *et al.* (2015) Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* 348, 921–925

18 Lieber, M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79, 181–211

19 Liang, F. *et al.* (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5172–5177

20 Daley, J.M. *et al.* (2005) Nonhomologous end joining in yeast. *Annu. Rev. Genet.* 39, 431–451

21 Emerson, C.H. and Bertuch, A.A. (2016) Consider the workhorse: Nonhomologous end-joining in budding yeast. *Biochem. Cell Biol.* 94, 396–406

22 Shibata, A. *et al.* (2011) Factors determining DNA double-strand break repair pathway choice in G2 phase. *EMBO J.* 30, 1079–1092

23 Shahar, O.D. *et al.* (2012) Live imaging of induced and controlled DNA double-strand break formation reveals extremely low repair by homologous recombination in human cells. *Oncogene* 31, 3495–3504

24 Dickerson, J.E. *et al.* (2011) Defining the role of essential genes in human disease. *PLoS One* 6,

e27368

25 Park, D. *et al.* (2008) Analysis of human disease genes in the context of gene essentiality. *Genomics* 92, 414–418

26 Kachroo, A.H. *et al.* (2017) Systematic bacterialization of yeast genes identifies a near-universally swappable pathway. *Elife* 6,

27 Yang, F. *et al.* (2017) Identifying pathogenicity of human variants via paralog-based yeast complementation. *PLoS Genet.* 13, e1006779

28 Hamza, A. *et al.* (2015) Complementation of Yeast Genes with Human Genes as an Experimental Platform for Functional Testing of Human Genetic Variants. *Genetics* 201, 1263–1274

29 Hamza, A. *et al.* (2020) Cross-Species Complementation of Nonessential Yeast Genes Establishes Platforms for Testing Inhibitors of Human Proteins. *Genetics* 214, 735–747

30 Sun, S. *et al.* An extended set of yeast-based functional assays accurately identifies human disease mutations. , *Genome Research*, 26. (2016) , 670–680

31 Laurent, J.M. *et al.* (2016) Efforts to make and apply humanized yeast. *Brief. Funct. Genomics* 15, 155–163

32 Kuzmin, E. *et al.* (2018) Systematic analysis of complex genetic interactions. *Science* 360,

33 Tong, A.H.Y. and Boone, C. Synthetic Genetic Array Analysis in Saccharomyces cerevisiae. , *Yeast Protocols*. (2005) , 171–192

34 Suzuki, Y. *et al.* (2011) Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. *Nat. Methods* 8, 159–164

35 Truong, D.M. and Boeke, J.D. (2017) Resetting the Yeast Epigenome with Human Nucleosomes. *Cell* 171,

36 Haase, M.A.B. *et al.* (2019) Superloser: A Plasmid Shuffling Vector for *Saccharomyces cerevisiae* with Exceedingly Low Background. *G3: Genes|Genomes|Genetics* 9,

37 Agmon, N. *et al.* (2020) Phylogenetic debugging of a complete human biosynthetic pathway

transplanted into yeast. *Nucleic Acids Res.* 48, 486–499

38 Botstein, D. and Fink, G.R. (2011) Yeast: an experimental organism for 21st Century biology. *Genetics* 189, 695–704

39 Winzeler, E.A. *et al.* (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* 285, 901–906

40 Li, Z. *et al.* (2011) Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat. Biotechnol.* 29, 361–367

41 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126

42 Mitchell, L.A. *et al.* (2017) Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* 355,

43 Papapetrou, E.P. and Schambach, A. (2016) Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol. Ther.* 24, 678–684

44 Lee, M.E. *et al.* (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* 4, 975–986

45 Doench, J.G. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191

46 Kearse, M. *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649

47 Murata, S. *et al.* (2009) Molecular mechanisms of proteasome assembly. *Nat. Rev. Mol. Cell Biol.* 10, 104–115

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

49 Beck, F. *et al.* (2012) Near-atomic resolution structural model of the yeast 26S proteasome. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14870–14875

50 Verwaal, R. *et al.* (2007) High-Level Production of Beta-Carotene in Saccharomyces cerevisiae by Successive Transformation with Carotenogenic Genes from Xanthophyllomyces dendrorhous. *Appl. Environ. Microbiol.* 73,

51 Chen, Y. *et al.* (2016) Lycopene overproduction in Saccharomyces cerevisiae through combining pathway engineering with host engineering. *Microb. Cell Fact.* 15,

52 Bahieldin, A. *et al.* (2014) Efficient production of lycopene in Saccharomyces cerevisiae by expression of synthetic crt genes from a plasmid harboring the ADH2 promoter. *Plasmid* 72,

53 Rébora, K. *et al.* (2001) Yeast AMP pathway genes respond to adenine through regulated synthesis of a metabolic intermediate. *Mol. Cell. Biol.* 21, 7901–7912

54 MGC Project Team *et al.* (2009) The completion of the Mammalian Gene Collection (MGC). *Genome Res.* 19, 2324–2333

55 Lamesch, P. *et al.* (2007) hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics* 89, 307–315

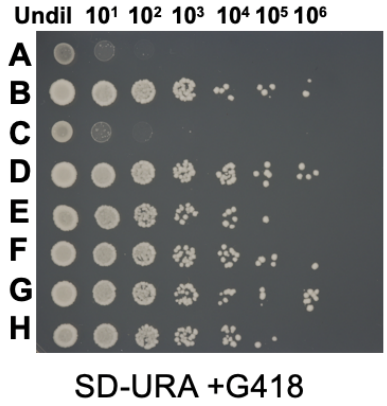
56 Daniel, J.A. *et al.* Eliminating Gene Conversion Improves High-Throughput Genetics in Saccharomyces cerevisiae. , *Genetics*, 172. (2006) , 709–711

57 Kramer, K.M. *et al.* (1994) Two different types of double-strand breaks in Saccharomyces cerevisiae are repaired by similar RAD52-independent, nonhomologous recombination events. *Mol. Cell. Biol.* 14, 1293–1301

Supplementary Information

Figure S1

A



ROW	MATING EXPERIMENTS
A	MATa-URA X MATa-G418 (- control)
B	MATa-URA X MATαG4-18 (+ control)
C	MATa-URA X MATα-URA (- control)
D	MATa-G418 X MATα-URA (+ control)
E	*MATa-511b::CrtE*- G418 X MATα-URA
F	MATa-*USERX-1*::*CrtB*- G418 X MATα-URA
G	MATa-*FgF20*::*Crt*I- G418 X MATα-URA
H	MATa-*FgF24*::*Crt*Y- G418 X MATα-URA

B

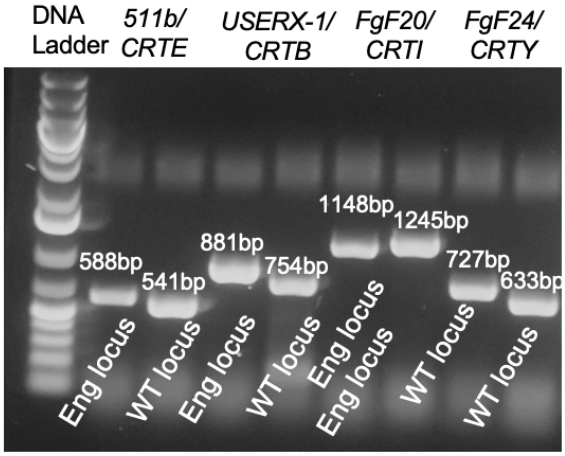


Figure S1. Generation of individual heterozygous diploids of each carotenoid pathway genes

A. SGA- Mat-A haploid strains harbouring each carotenoid gene individually were mated with the opposite SGA Mat- α haploid wild type strains to create heterozygous diploids [Refer to Methods and Materials, section 7.4]. Negative controls, showed no growth on double selection due to same mating type (no-mating) (row A) and same selection (row C). Positive controls showed growth due to the double selection (using plasmids with two distinct selections) and efficient mating. **B.** Locus specific PCR confirmation shows both engineered (carotenoid genes) and wildtype sequences (safe-harbour sites) are present.

Figure S2

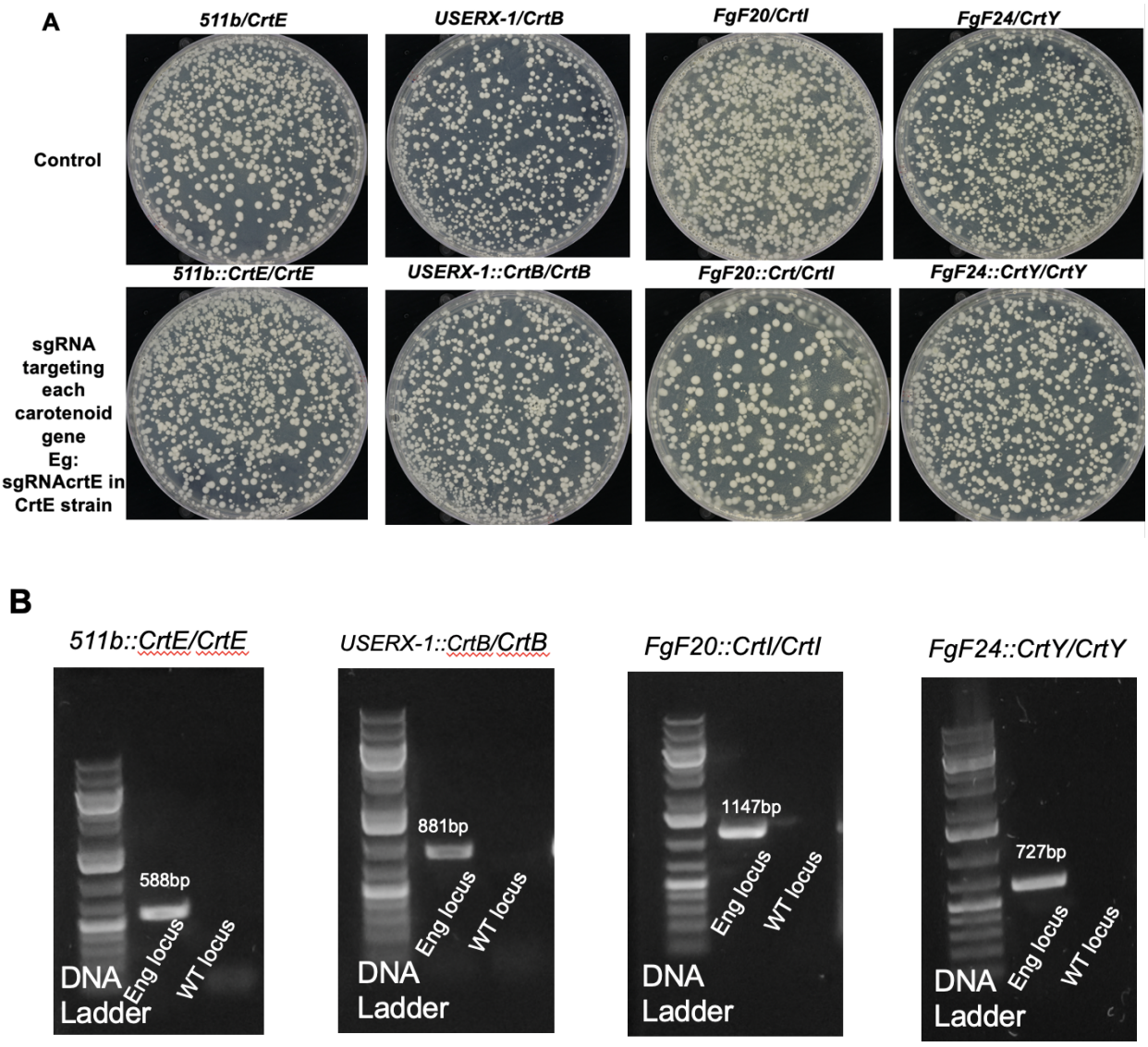


Figure S2- Using CRISPR-Cas9 to create individual homozygous diploids for the carotenoid pathway genes

A. Each heterozygous carotenoid strain was transformed with a control plasmid (pCRISPR-Cas9) and also with an sgRNA targeting the specific safe harbour site that was modified *[Refer to Methods and Materials, section 7.2*]. For example, the heterozygous strain carrying the 5*11b* wildtype sequence on one allele and the *CrtE* gene on the other allele was transformed with the sgRNA targeting the wildtype *511b* sequence. **B**. Colonies from experimental plates from A (treated with Cas9 and a sgRNA) were genotyped, and showed the conversion of the wild type locus to its engineered counterpart.