Title
Iterative library generation and FACS screening for increased production of the pharmaceutical precursor L-DOPA in yeast

Author
Savitskaya, Judith

Publication Date
2019

Peer reviewed|Thesis/dissertation

Iterative Library Generation and FACS Screening for Increased Production of the Pharmaceutical Precursor L-DOPA in Yeast

by

Judith Savitskaya

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John Dueber, Co-Chair
Professor Adam Arkin, Co-Chair
Professor Hana El-Samad
Professor Jeremy Thorner

Spring 2019

ABSTRACT

Iterative library generation and FACS screening for increased production of the pharmaceutical precursor L-DOPA in yeast

by

Judy Savitskaya

Joint Doctor of Philosophy
with University of California, San Francisco
in Bioengineering

University of California, Berkeley

Professor John Dueber, Co-Chair
Professor Adam Arkin, Co-Chair

Optimizing microbial hosts for large-scale production of valuable metabolites has two main challenges: (i) maximizing the expression and function of the proteins needed to perform the desired reactions; and, (ii) adapting host metabolism to support these new reactions and remove unwanted or toxic side products. Although considerable effort has been focused on the first challenge, the methodology developed and described in this dissertation addresses the second challenge. As a test case, I sought to increase production of L-DOPA, a pharmaceutically-relevant metabolite and precursor to the benzylisoquinoline (BIA) class of medicinal molecules, in the yeast *S. cerevisiae*. Production of L-DOPA and derived BIAs in yeast can be accomplished through the action of CYP76AD1, a heterologously-expressed cytochrome P450 enzyme from beet. Hence, I sought to devise strategies for isolating strain variants that carry genome modifications that improve CYP76AD1-dependent L-DOPA production. My approach was based on the assumption that endogenous factors in yeast restrain or impede this process. To perform multiple rounds of mutagenesis and screening, we constructed an *in vitro* barcoded transposon-disruption library. This library was introduced and integrated into the host genome by homologous recombination. The resulting variants were screened using a biosensor in which L-DOPA produced by the cells is converted to a fluorescent derivative via the action of the enzyme DOPA dioxygenase (DOD). Thus, I was able to use high-throughput fluorescence-activated cell sorting (FACS) to enrich for the desired variants. I conducted this approach, iteratively, for three rounds, i.e. improved strains obtained from the preceding rounds were transformed with the transposon-disruption library and re-screened. In the first two rounds of screening, I identified deletions that improved biosensor compartmentalization and, consequently, improved the reliability of the read-out for L-DOPA production. In the final round, I discovered that deletion of the gene encoding a heme oxygenase (*HMX1*) that is

localized to the endoplasmic reticulum and involved in heme degradation increased both total cellular heme content and L-DOPA production (as monitored by measuring its derivative dopamine as a proxy). I demonstrated further that deleting *HMX1* may represent a general strategy for improving the performance of heterologous P450 enzymes in yeast because the absence of Hmx1 also enhanced the ability of a second P450 enzyme, BM3, to generate its product.

ACKNOWLEDGEMENTS

As the wise Dave Matthews teaches us: “it’s not where you are, but who you’re with that really matters”. This PhD was built on the support of my advisors, labmates, friends, and family. I’ll give short mentions here, but the amount of appreciation and affection I feel for each of these people could fill an entire dissertation by itself.

Thank you to John Dueber, who dealt with my ambition and distractibility by looking for the lessons in our tribulations and taking advantage of teachable moments. John, I’ve learned more from you (despite some reluctance) than you would ever guess.

Thank you to Adam Arkin, who would happily embark on ethereal thought journeys with me and challenge my intellect at every turn. Every time I came to Adam with my brain in a cloud of confusion, he would find the nucleating concept that would crystallize my cloud into an orderly set of next steps. One of my biggest regrets will always be that I didn’t fully take advantage of the opportunity he presented to pursue the craziest ideas out there.

Day after day, well after well of pipetting, I felt at home in lab with my labmates nearby. To Will, Mike, Zach, and Luke: you built the Dueber lab and your legacy lives on in all of our projects, especially mine. You have had an outsized influence on how each lab member develops as a scientist – even new members who you have never met. Francesca was the undergraduate mentee that grad students dream of. Watch out for this one, she’s going somewhere big. Jen is my confidant, my adventure partner, and my pub buddy. Harneet is probably the coolest person I know. I’m incredibly thankful to have spent my entire PhD journey with Tammy. After all, who else would blast a Sia playlist with me while finishing up some midnight cloning? Shakked is always ready to try something new or take a chance, be it traveling or playing a game or exploring an LED forest. For Andrew, I’ll get up for 7am yoga – it’s worth it for everything I learn from him on the drive and in our climbing session after. Lastly, Parry was my partner in thought, in action, and in mentoring undergraduates. Parry is a force of nature, I trusted his experimental results like no one else’s because he only executes when all of the pieces of the experiment are clean, in place, and perfectly controlled. The way our brains would sync in subgroup meeting, wordlessly exchanging ideas and weeks’ worth of experimental plans, was an ethereal experience. Another enormous regret of mine is that our projects diverged and our partnership with them. I’m lucky that the Bay Area is so sticky and we’ll both stay around – hopefully we can work together again in the future!

To my family, thank you for learning not to ask about my graduation date! You made holidays, trips, and phone calls into fun, love-filled intermissions in my lengthy PhD journey. Mom and Borya, every phone call with you ends in "and if you flunk out of your PhD, you know you can always just come live with us in New Jersey!" I know, I know. And thank you, but I think I'll just graduate. Still, you'd be surprised to learn how much solace your unappealing offer provides me. Luba, Sasha, Vadim, Sonja, and Jane: you've made California into a true home for me. I feel completely enveloped by our family here and find so much joy in our hilarious conversations and attempts to make Sasha stop talking. You've changed the holidays from a time for travel into a time for tradition, Chinese food, and "oohing and ahhing" at children's gifts. Lastly, thank you for helping me teach Derek Russian. He'll be fluent one day, but probably only in the vocabulary required to re-tell Sasha's stories.

Derek, who are you? I'm still learning every day. I know you're an instigator (…mostly of adventure…), a therapist, a repairman, a Legend of Zelda character, a master scheduler, a cheese eater, a professional gear shopper, a google sheets expert, a list-making extraordinaire, a people-hating people-lover, and a pessimistic optimist. What else? I think figuring it out will be my next big research project – let's get started!

TABLE OF CONTENTS

1 Introduction

A large variety of industrially important molecules are produced through metabolic engineering of fermentative organisms. One of the major challenges in engineering these manufacturing systems is optimization of the host organism's genome. This includes increasing production of source metabolites, removing cellular processes that impair the heterologous pathway, and adjusting the cellular environment to support chemistries performed by new enzymes. In lieu of a robust *in silico* approach for predicting beneficial genomic mutations, three experimental strategies are available. 1) Engineers can rationally choose genomic targets for mutations based on pathway information in databases like Metacyc. In this approach, only a limited number of variants can be tested, and the best mutations may not be easy to predict. If a screen or selection is available, other options are to 2) mutagenize the strain or 3) search for improvements within existing natural variation. For mutagenesis, the mutant pool would need to be very large to query every gene and most combinations, which may be challenging if using a screen. For natural variation approaches, exploration is limited to the genetic diversity already sampled by evolution. Furthermore, either a few improved strains can be analyzed by full genome sequencing or a pool of strains can be analyzed in bulk, at the loss of information regarding interactions between mutations.

Although random mutagenesis produces too many variants to screen, most of these mutations will be neutral. Two classes of mutations are much more likely than random mutagenesis to produce a phenotype: gene deletion and gene overexpression. Targeted screens of gene deletions and overexpressions offer the opportunity to both sample genome-wide modifications without prior knowledge, while remaining easy to interpret (Gorsich et al., 2006; Özaydın, Burd, Lee, & Keasling, 2013; H. H. Wang et al., 2009; Warner, Reeder, Karimpour-Fard, Woodruff, & Gill, 2010; Yoshikawa et al., 2011). While this approach limits the class of hits that may be found (ex: it will be unable to find mutants that break allosteric regulation or regulatory regions of DNA), it ensures a complete search for mutants whose effect is mediated through a change in gene expression. Moreover, deletion and overexpression are large changes that may substantially upset the balance of molecular species in the cell compared with most point mutants that do not destroy protein function. Lastly, such screens can leverage the wide array of experimental data available in the host organism; past studies are more likely to describe phenotypes found for full deletions or

overexpressions, while particular point mutations found through random mutagenesis are far less likely to appear in the existing literature.

In this dissertation, I report the results of a fluorescence-based screen for production of a high-value metabolite using three libraries 1) deletion collection, 2) a barcoded transposon-based disruption library, and 3) a gene overexpression library. The deletion collection was purchased (Dharmacon, Lafayette, CO). The disruption library was developed in collaboration with Ryan Protzko and will be described in Chapter 4. The overexpression library was constructed and characterized in this study and will be discussed in Chapter 6.

1.1 Screening by multiple rounds in metabolic engineering

A common strategy for optimizing a host strain for fermentative production is to create mutations in the host genome and screen or select for enhanced strains. Iterative rounds of mutation and screening/selection enable researchers to find multiple mutations that act synergistically to improve the host strain. However, iterative rounds of screening or selection often require reconstruction of new libraries of mutated strains. Because of this challenge, metabolic engineering studies often consist of only a single round of mutagenesis and screening.

We created a barcoded transposon-mediated gene disruption library in an *in vitro* yeast genomic DNA preparation. Using Randomly Barcoded Transposon Sequencing (RB-TNSEQ), barcodes in each transposon can be uniquely mapped to the genomic integration site of the transposon (Wetmore et al., 2015). The key advantage of this strategy is that the transposon-disrupted genome fragment library needs to be generated once and the barcode-insertion position associations only need to be characterized once. The same library can be integrated with high efficiency into any *Saccharomyces cerevisiae* strain background by homologous recombination. Insertional mutations enriched through selection or screening can then be rapidly identified by PCR amplification and sequencing the associated barcode (BarSEQ). This enables iteration to identify a multi-insertion strain with improved performance.

The details of this construction and characterization process are described in Chapter 4.

1.2 Benzylisoquinoline alkaloid pathway

Benzylisoquinoline alkaloids (BIAs) comprise a pharmaceutically valuable class of natural molecules. The most well-known molecule in this class is the analgesic drug morphine, as well as its derivatives, codeine and oxycodone. However, there exists a wealth of natural BIAs and derivative ”unnatural” molecules that have a chemical structure suggestive of bioactivity, but have not been produced in sufficient quantities for pharmaceutical testing (Hagel & Facchini, 2013). Some examples include noscapine and sanguinarine. Since these molecules have multiple stereocenters, they are difficult to produce using chemical synthesis. Currently, BIAs commercially sold as drugs are extracted from their native plant producers, but engineering plants to produce novel molecules of interest is slow and laborious. Several academic labs and companies have proposed a biosynthetic route to these products in *S. cerevisiae* or *Escherichia coli* as a faster alternative (DeLoache et al., 2015; Galanie, Thodey, Trenchard, Filsinger Interrante, & Smolke, 2015). However, the product titers obtained thus far by these academic and industrial projects are insufficient to be economically viable.

The molecules in the BIA pathway are diverse, but they are all derived from a benzylisoquinoline structure, consisting of a benzene ring fused to a pyrimidine ring linked to a benzyl moiety. (S)-reticuline is a common precursor for the production of a wide variety of BIA molecules and has been used as an endpoint in metabolic engineering studies (Figure 1.1C). All BIAs are derived from 2 units of L-tyrosine, beginning with the condensation of two tyrosine derivatives, dopamine and 4-HPAA (Figure 1.1B). Tyrosine is present at very low concentrations in *S. cerevisiae*. The pathway contains several branch points and feedback mechanisms (Figure 1.1A).

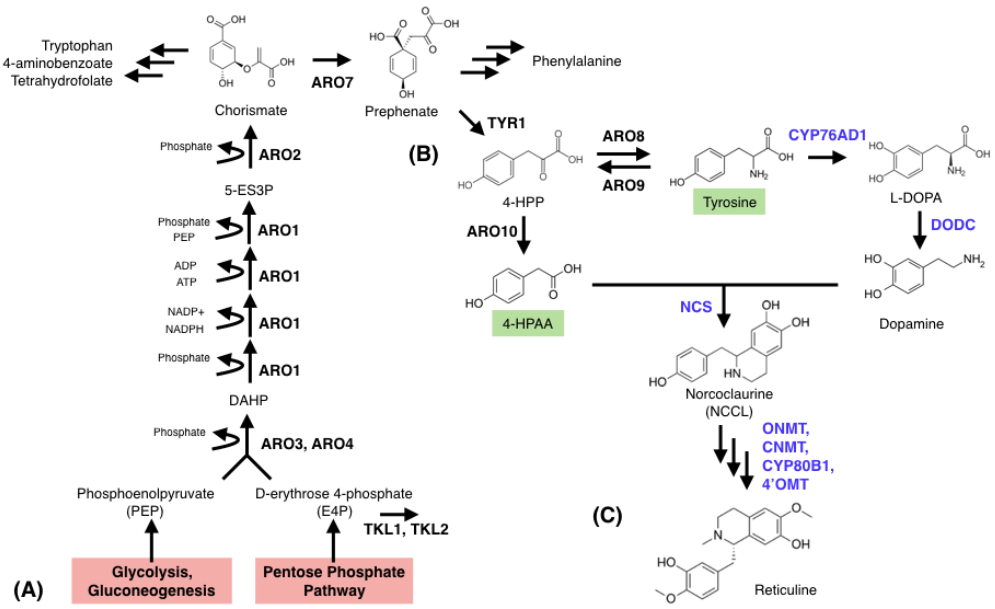

Figure 1.1 Tyrosine, produced via the aromatic amino acid pathway, is used for heterologous production of BIAs. Native genes are shown in black bold and heterologous genes are shown in blue bold. Green boxes mark the native molecules used as substrates in the BIA production pathway. A) The aromatic amino acid pathway uses products of the pentose phosphate pathway and glycolysis to produce tyrosine, tryptophan, and phenylalanine. B) TYR1 represents the first committed step to tyrosine production, 4-HPP is converted to either tyrosine or 4-HPAA. As part of the heterologous pathway, tyrosine is oxidized and decarboxylated to dopamine, then condensed with 4-HPAA to form the first BIA, norcoclaurine. C) Norcoclaurine is converted into central BIA product, (S)-reticuline via four enzymatic steps.

DeLoache *et al*. recently described an enzyme that catalyzes the conversion of tyrosine to L-DOPA, the only previously uncharacterized step in heterologous BIA synthesis in *S. cerevisiae* (DeLoache et al., 2015). They accomplished this by finding an enzyme, DOPA deoxygenase (DOD), which converts L-DOPA into a group of fluorescent molecules collectively called betaxanthins (Figure 1.2). They then integrated this enzyme into the yeast genome and screened a library of tyrosine hydroxylase (CYP76AD1) mutants for high betaxanthin production. Using this “betaxanthin biosensor”, they showed that the betaxanthin fluorescence was closely correlated with the L-DOPA levels in a few mutants selected for analysis by LC/MS. With an improved tyrosine hydroxylase and a set of six heterologous enzymes

described in previous papers, they developed a yeast strain that could produce (S)-reticuline.

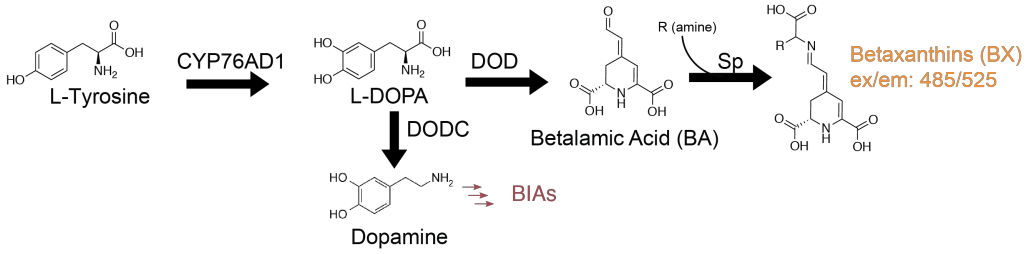

Figure 1.2 Betaxanthin constitutes a biosensor for L-DOPA. The heterologous pathway to L-DOPA includes the enzyme CYP76AD1, which oxidizes native tyrosine to L-DOPA. Heterologous enzyme DOD can convert L-DOPA into betalamic acid, which then spontaneously condenses with native amines to form betaxanthins. Expression of heterologous DODC decarboxylates L-DOPA to dopamine, which can be used in downstream reactions to produce BIAs.

In previous studies, titer of final BIA molecules like (S)-reticuline and morphine were less than 300µg/L (DeLoache et al., 2015; Galanie et al., 2015). For comparison, the standard dosage for morphine is between 5 and 15 mg. Titers would have to be increased in order to make fermentative production of these molecules commercially viable. I found that low L-DOPA concentration at the entry of the BIA pathway is a bottleneck in increasing production of valuable downstream chemicals. Figure 1.3 demonstrates that (S)-reticuline production in our engineered strain scales linearly with fed L-DOPA. This indicates that the pathway is substrate-limited and increased L-DOPA production would directly lead to increased (S)-reticuline production.

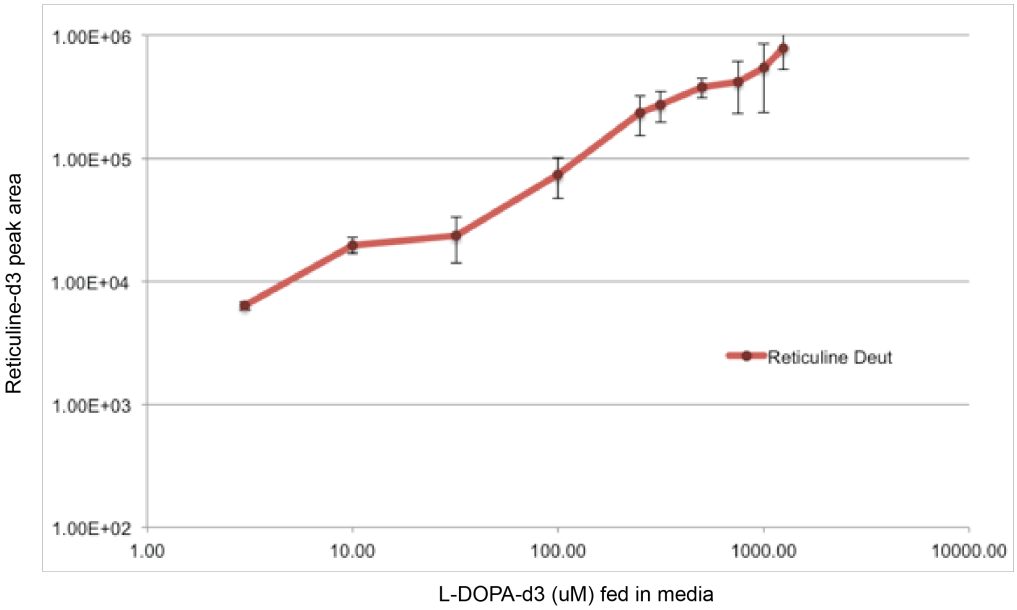

Figure 1.3 L-DOPA is limiting in the production of downstream product (S)-reticuline. (S)-reticuline can be produced from tyrosine via seven enzymatic steps, the first two of which are CYP76AD1 and DODC to produce dopamine. Increasing concentrations of deuterated L-DOPA fed in culture media are converted into increasing amounts of deuterated (S)-reticuline. (S)-reticuline measurements were made after 48 hours of growth. At the time, no (S)-reticuline standard was available so we were unable to make a standard curve.

Other groups have made modest improvements to tyrosine production using rationally chosen genomic modifications to native yeast enzymes, with the goal of producing BIA molecules downstream. However, these approaches have only yielded moderate success. Christina Smolke's lab, for example, achieved only a 1.6X increase in (S)-reticuline production (12.3µg/L to 20µg/L) by overexpression three genes: 1) ARO4-fbr, 2) ARO7-fbr, and 3)Tkl1p (Galanie et al., 2015). Similarly, Vincent Martin's lab described an modest increase in tyrosine production by creating five strain modifications: 1) ARO4-fbr, 2) ARO7-fbr, 3) TyrC) from *Z. mobilis*, 4) Δ*aro10*, and 5) Δ*zwf1* (Gold et al., 2015).

While these two rational engineering approaches produced modest improvements in tyrosine titer, we hypothesized that we could learn more from an unbiased genome-wide engineering approach. Importantly, some of the most beneficial mutants may be those that support CYP76AD1 function since this enzyme is a cytochrome P450 that catalyzes a redox reaction, described in Section 1.3. Several studies have successfully used genome-

wide libraries, like the deletion collection, to identify useful mutations in a variety of pathways (Özaydın et al., 2013; H. H. Wang et al., 2009; Warner et al., 2010; Yoshikawa et al., 2011). As described in Section 1.1, a novel aspect of the genome-wide screen described in this dissertation is the use of iterative rounds of mutagenesis and screening.

1.3 P450 enzymes

Endogenous pathways for producing BIAs includes several cytochrome P450 enzymes (P450s), which are important in many other biosynthetic pathways to valuable pharmaceutical molecules such as artemisinin and paclitaxil or pigments such as betalain (Dietrich et al., 2009; Grewal, Modavi, Russ, Harris, & Dueber, 2018; Hagel & Facchini, 2013; Hawkins & Smolke, 2008; Jensen & Møller, 2010; Jung, Lauchli, & Arnold, 2011; Minami et al., 2008; Rasool & Mohamed, 2015; A. Stierle, Strobel, & Stierle, 1993; Withers & Keasling, 2007). Many efforts to express P450s in fermentative host organisms have been successful, but challenged by difficulties in the expression, folding, cofactor usage, and protein-protein interactions of P450s in their non-native hosts (Jung et al., 2011; Michener, Nielsen, & Smolke, 2012; Whitehouse et al., 2009). P450 enzymes require a reductase partner that can pass electrons from NAD(P)H cofactors. They also must be loaded with an iron-containing heme group. Natively, P450s are anchored in the endoplasmic reticulum via a transmembrane domain. Each of these factors contributes to the challenge of producing high-titer P450 products in microbial hosts, especially when the biosynthetic pathways contain multiple P450 enzymes. While heterologous P450 expression and activity have been shown in *E. coli*, with a few exceptions (Biggs et al., 2016; Edgar, Li, Qiao, Weng, & Stephanopoulos, 2016) such work has required modifications to the transmembrane domain of the P450 and fusion of a reductase partner to overcome the challenges of using a prokaryotic host (Ajikumar et al., 2010; Chang, Eachus, Trieu, Ro, & Keasling, 2007; Leonard & Koffas, 2007; Renault, Bassard, Hamberger, & Werck-Reichhart, 2014). On the other hand, *S. cerevisiae* have an endoplasmic reticulum and a native reductase, which make this organism a suitable host for P450 expression.

Previous studies have used protein engineering and directed evolution to modify P450 performance through altered substrate specificity, improved P450-NAD(P)H coupling efficiency, increased tolerance to new reaction conditions, and even altered chemical catalysis (Chen, Huang, Kan, Zhang, & Arnold, 2018; Dietrich et al., 2009; J. C. Lewis et al., 2010; McIntosh, Farwell, & Arnold, 2014; Whitehouse et al., 2009). However, few studies have

investigated the optimization of the cellular environment in which heterologously expressed P450s are expressed. Targeting modifications to the host genome may be a promising strategy for improving P450 function since these enzymes use host resources for expression, folding, and cofactor loading.

2 Using FACS-based screening to identify increased bioproduction of fluorescent molecules

2.1 Introduction to sorting methodology

Fluorescence-activated cell sorting (FACS) is a method used to enrich for a subpopulation of cells that contains the fluorescent marker of interest to the study. A cell culture is placed into the FACS instrument, which then uses negative pressure to pull the culture up into its fluidics lines. The culture travels through a "sorting chip" in which it is mixed with sheath fluid (usually PBS) and ejected out of a "droplet nozzle". An acoustic wave generated by a piezoelectric device enables the formation of distinct, stable droplets at a high speed. Immediately before the droplets are released, the cells pass through the excitation laser in the FACS instrument. Photons excited by the laser emit light that passes through a filter set and is ultimately measured at six different photo-detectors. These measurements are recorded for each "event" which is an instance of the lasers' light transmission decreasing as an object passes through. Events roughly correspond to cells, though aggregates or particulate matter can appear as events as well.

When sorting, the user draws "gates" to specify the fluorescence conditions of the cells he or she is interested in enriching for. The detection of an event that satisfies the gate definitions is recorded and the stream is charged at the moment that the droplet containing that event exits the nozzle. The instrument can apply a positive or negative charge or allow the droplet to remain neutral. The charged or neutral droplets then pass through two electrode plates (positive and negative) that deflect the droplets. This process generates three distinct streams: left, right, and center, where center is usually diverted to waste.

A common use of FACS is to label cells with fluorescent antibodies targeted at different cell-surface markers, then sort the cells to enrich for cells of a given type or that have a specific cell surface protein. The majority of common uses for FACS are based on on-off- switches, that is, the cells are either fluorescent in the channel of interest or not. In this work, I used FACS to discriminate different fluorescence intensities to infer the amount of a fluorescent molecule present inside the yeast cell, which is an uncommon use of this technology. The minority of past studies have used sorting to bin fluorescence intensities rather than simply the presence or absence of fluorescence (Cambray, Guimaraes, & Arkin, 2018; Coradetti et al., 2018; Goodman, Church, & Kosuri, 2013).

The goal of this study was to identify mutants that have higher production of a fluorescent molecule, betaxanthin, as a proxy for increased flux through the pathway to L-DOPA. Throughout this dissertation, I reference two sorting techniques, which I named “threshold sorting” and “sort-seq”. I used both methodologies in each round of mutation and sorting, but I found that the threshold sorting produced robust mutants, while sort-seq generated almost exclusively false positives.

In threshold sorting, I simply sorted for the top 0.5% fluorescent cells in any library. I did this by flowing a mixed population of cells through the flow cytometer, setting a sorting gate at the top 0.5% and collecting sorted cells (Figure 2.1A). The mutations in the sorted cells could be identified by sequencing the DNA barcode that was inserted into the mutant either in the commercially available deletion collection or in the transposon-based disruption library that I built with Ryan Protzko.

In the sort-seq methodology, I sorted the top 5% most fluorescent cells in the library populations into a bin and sequenced the DNA barcodes of these mutants in a pooled sequencing assay. As a control for growth and stresses induced by the sorting process, I maintained an “unsorted” control population that was run through the sorter with no gate drawn (cells of any fluorescence were selected) and sequenced using the same methodology. In order to infer the number of copies of each mutant in the sorted and unsorted pools, I counted the number of instances of each DNA barcode and matched it to the mutant it represents (Figure 2.1B).

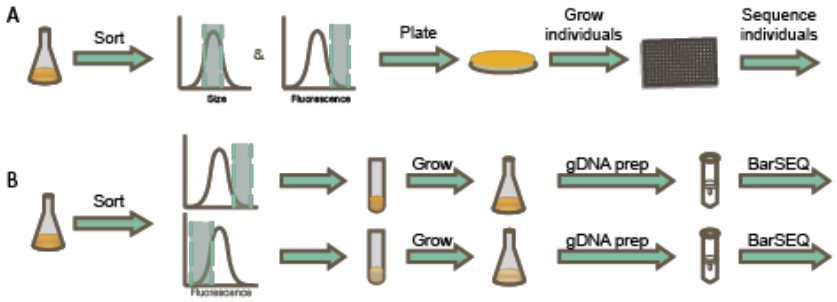

Figure 2.1 Schematics of sorting methodologies. A) In “threshold sorting”, cells are sorted for the 35-65th percentile of size (to capture 30% of the population), as measured by forward scatter area in the flow cytometer. Size-selected cells are then subsorted for high fluorescence (top 0.5-1% of population). Sorted cells are plated on solid media, then individual colonies are grown in culture media. Barcode sequencing determines the identity of the most fluorescent variants. B) In “sort-seq”, cells are sorted for the 35-65th percentile of size (to capture 30% of the

population), as measured by forward scatter area in the flow cytometer (not shown). Size-selected cells are sorted for high and low fluorescence. 800,000 cells below the 5th percentile of fluorescence and 800,000 cells above the 95th percentile of fluorescence are collected and grown to dense culture. Genomic DNA is prepared from these cultures and barcode sequencing identifies the deletions present in these two populations.

2.2 A mutant with 2x greater activity can be enriched through sorting

I ran experiments to verify that FACS methodology could be used to discriminate between mutants with different amounts of the fluorescent molecule, betaxanthin. Prior work from DeLoache *et al* demonstrated that two single nucleotide mutations in the enzyme CYP76AD1 (W13L and F309L) increased betaxanthin production in a yeast host and therefore increased cell fluorescence (DeLoache et al., 2015). This publication shows that the mutations had an additive effect through a liquid culture experiment, measuring with a plate reader. Figure 2.2 is reproduced with permission from DeLoache *et al* to demonstrate this.

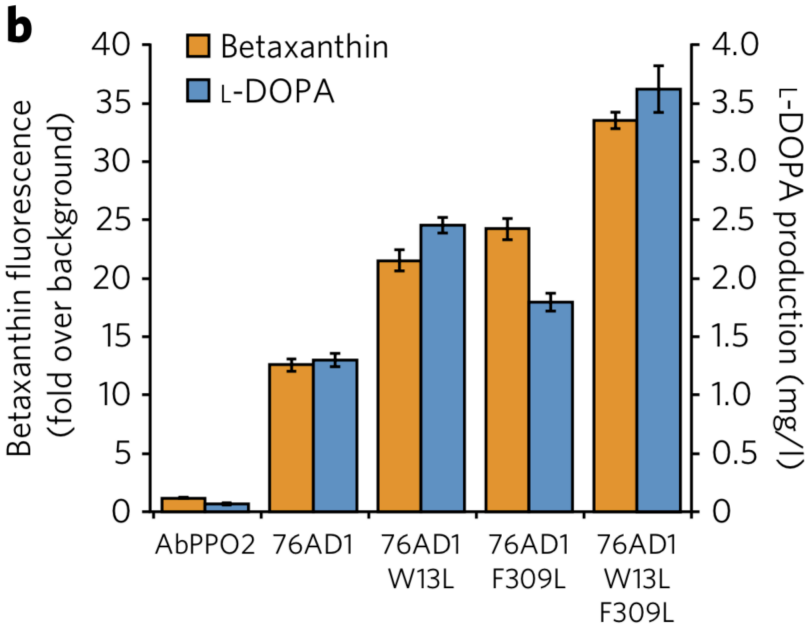

Figure 2.2 CYP76AD1 variants increase betaxanthin and L-DOPA production. Two mutations in CYP76AD1, W13L and F309L, each increase betaxanthin production by approximately 2-fold compared with wild type CYP76AD1. Together, these mutants have an additive effect on

betaxanthin fluorescence. Figure reproduced with permission (DeLoache et al., 2015).

To develop the sort-seq methodology described in section 2.1, I created an artificial “library” by mixing cultures of yeast containing two of the enzyme variants described in Figure 2.2 and sorted using different conditions. For these experiments, I also engineered the highest producing strain (containing the CYP76AD1 W13L F309L variant) to express a red fluorescent protein so that I could identify the strain in a mixture with other variants. I sorted with a gate drawn at the top 5% of the population in the betaxanthin emission channel, plated the sorted cells on solid agar, then tested the quality of the sort by counting the percent of colonies that expressed mKate2. Sorting can be performed with different stringency criteria, which dictate whether droplets containing multiple cells are sorted. Since the higher producing strains are a small percentage of the total population, sorting using stringent purity criteria would take a long time in order to enrich for enough cell mass to perform sequencing. Instead, I found that enriching the population through less stringent “yield” sorting, followed by more stringent “purity” sorting enabled me to enrich for the higher betaxanthin producing cells much faster than “purity” sorting alone. Figure 2.3 shows the results of this experiment.

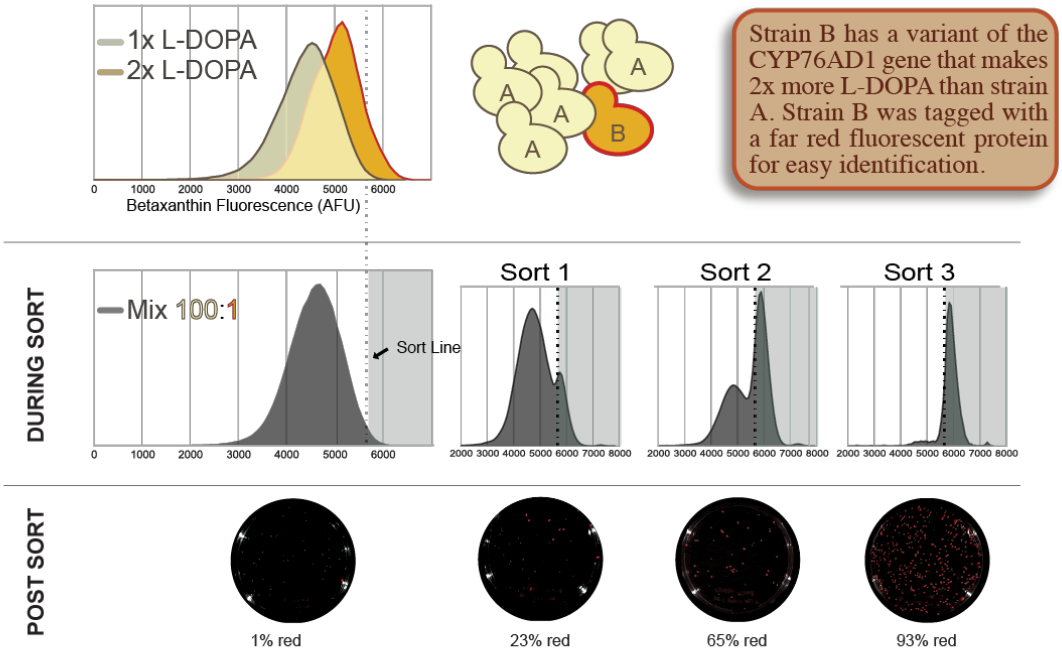

Figure 2.3 Multiple rounds of sorting with increasing stringency enable enrichment of a low-frequency variant that produces 2-fold more L-DOPA than the high-frequency variant. "B" cells contain the W13L mutation in CYP76AD1, making them produce more betaxanthin than "A" cells, which have wild type CYP76AD1. These two cell types were mixed at a 100:1 ratio of A:B cells. B cells were labeled with mKate (red fluorescent protein) to allow for easy identification post-sort. Multiple sorts enabled enrichment of the B cells to 93% of the population.

3 Screening a pooled deletion collection for mutants with increased bioproduction of a fluorescent molecule

3.1 Pooling the arrayed yeast deletion collection

The yeast deletion collection is commercially sold in a format containing 53 96-well plates of strains suspended in glycerol and frozen (Brachmann et al., 1998) (Dharmacon, Lafayette, CO). In order to perform a single transformation to integrate the betaxanthin biosensor genes, the individual strains needed to be pooled into one culture. Using robotic liquid handling (High-throughput sequencing center, UC Berkeley), I generated a pooled culture of the strains in the deletion collection. This pooled culture was distributed into 2mL aliquots and frozen at -80°C for use in future experiments.

For the first round of screening we integrated genes comprising the betaxanthin biosensor into a pooled culture of all strains arrayed in the yeast deletion collection (Brachmann et al., 1998). Barcode sequencing (BarSEQ) demonstrated acceptable uniformity of coverage for the 4,785 unique ORF deletions reported to be present in the deletion collection (Figure 3.1) (Smith et al., 2009). In Figure 3.1, the read counts for 4785 ORFs were binned and plotted as a histogram. For each ORF, the read counts for the UPTAG and DNTAG were both recorded, and the maximum of these two values is shown here. This is to account for mutations or poor sequencing for some barcodes. 510 strains (10.7% of the library) did not have any UPTAG or DNTAG read counts, which is consistent with previous re-sequencing effort (Eason et al., 2004). The median read count is 2,636. A majority of the strains (63.6%) have a read count within a factor of 2 (1,318) or a multiple of 2 (5,272) of the median count.

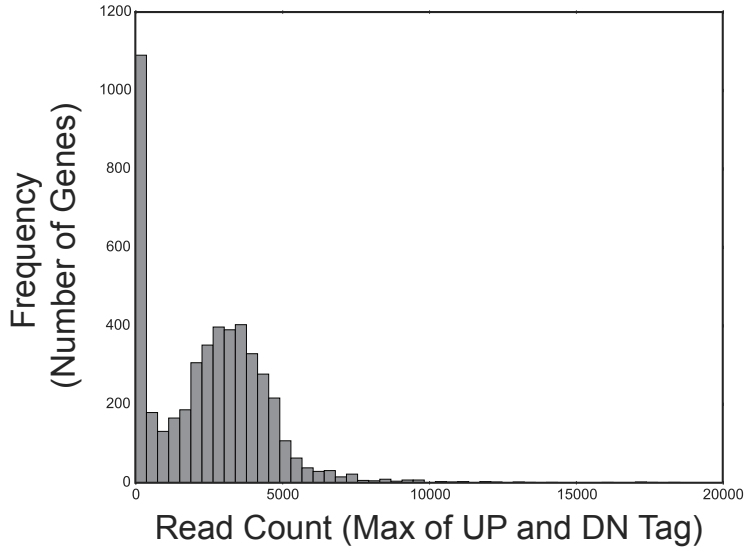

Figure 3.1 Barcode read count distribution for pooled deletion collection. A histogram for the number of reads for each barcode reveals that many barcodes were at a low frequency or absent from the pooled deletion collection. Of the barcodes present, most clustered around 4000 reads per barcode, with very few at >10,000 counts. This suggests that the variants in the library are fairly evenly distributed and no single variant is likely to overtake the population.

3.2 Threshold sorting of the deletion collection

Pooled transformants were screened for high betaxanthin production using fluorescence-activated cell sorting (FACS), gated for the top 0.5% of events. Individual sorted strains were grown on solid agar media (Figure 3.2). Out of approximately 1000 colonies, 30 were chosen by visual screening of colonies for high fluorescence and sequenced. Sequencing revealed that of these 30 colonies, 17 harbored a deletion of PDR8, a transcriptional activator of ATP-binding cassette (ABC) transporters and other drug resistance genes. Two of the 30 strains harbored a deletion of QDR2 and one of the 30 strains had IMA5 deleted. QDR2 and IMA5 are ABC transporters that are regulatory targets of PDR8. We hypothesized that these transporters are responsible for efflux of betaxanthin and their deletion increases betaxanthin retention, resulting in higher betaxanthin signals in single-cell assays like flow cytometry. We confirmed the phenotypes of these gene deletions by building markerless knockouts of PDR8, QDR2, and IMA5 in wildtype strains not from the deletion collection (M. E. Lee, DeLoache, Cervantes, & Dueber, 2015). In

these strains, liquid culture measurements of intra- versus extracellular fluorescence showed that deleting QDR2 dramatically increased compartmentalization of betaxanthin and accounted completely for the effect from the PDR8 deletion. Figure 3.2A shows the fluorescence distribution in flow cytometry and Figure 3.2B shows the fraction of the fluorescence that we found to be intracellular. In Figure 3.2B, the y-axis measures the intracellular fluorescence divided by the sum of the intracellular and extracellular fluorescence. The reason for IMA5 enrichment in the original screen is unclear, since its deletion alone did not increase compartmentalization (Figure 3.3).

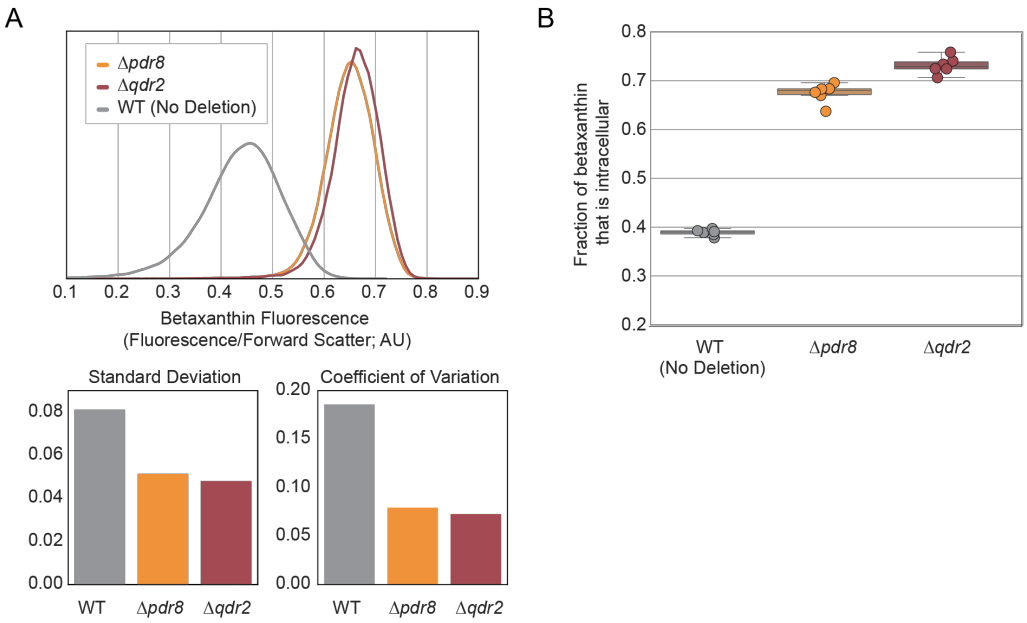

Figure 3.2 Initial screening identified deletions of PDR8 and QDR2 increase betaxanthin retention. Initial screening identified PDR8 and QDR2 as deletions that increase betaxanthin retention. A) Deletions of these two genes increased single cell fluorescence and decreased the variance in fluorescence measurements within an isoclonal population. The population standard deviation and coefficient of variation are shown below. The resulting coefficient of variation for the *Δpdr8* and *Δqdr2* strains was lower than the WT strain. B) In liquid media, intracellular betaxanthin retention was higher for *Δpdr8* and *Δqdr2* compared to WT. All strains include the same biosensor cassette. Six samples were measured; boxes represent the interquartile range (IQR; 25-75%). Whiskers represent the range of the data with the exception of outliers. Data points outside of the IQR +/- 1.5xIQR are considered outliers.

The remaining 10 colonies each contained a unique deletion and none of these deletions conferred increased fluorescence when the colonies were picked into liquid media and bulk fluorescence tested, suggesting they were false positives. Given that wild-type strains have high rates of betaxanthin export, we hypothesized that these colonies were positioned in close proximity to other colonies on the agar plate, increasing their perceived betaxanthin production.

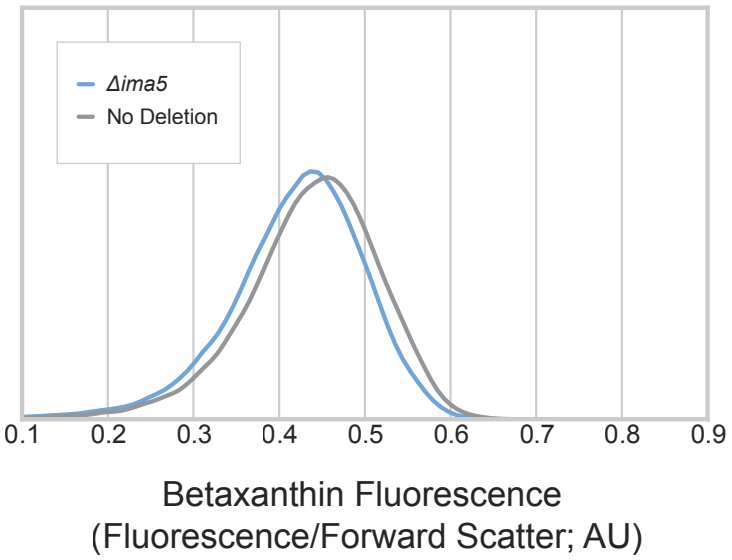

Figure 3.3 *Δima5* did not increase betaxanthin production. *Δima5* was enriched in the first round of sorting. Similar to QDR2, IMA5 encodes a multi-drug family transporter that is regulated by PDR8. However, reconstruction of the IMA5 deletion in a clean background strain shows no improvement in fluorescence over the control. We conclude that IMA5 was a false positive.

3.3 Sort-seq experiments on the pooled deletion collection

I also performed the sort-seq methodology described in Section 2.3 on the deletion collection library (containing the betaxanthin biosensor genes). Cells were sorted according to size and fluorescence, then sequenced at the barcode site as described in Section 2.3. Barcode read counts were recorded for three samples: 1) cells sorted by size and high fluorescence (called

"Top"), s) cells sorted by size and low fluorescence (called "Bottom"), and 3) cells sorted by size only (called "Control").

The purpose of the Control sample was to account for different growth rates of variants in the library, as described in Section 2.3. The raw read counts for each gene are displayed in Figure 3.4 (expressed as proportions of total read count). As expected, variants that were enriched in the Top sample are not enriched in the Bottom sample, and vice versa. This is a good indication that the variants enriched in each sample are not simply enriched due to a fitness advantage.

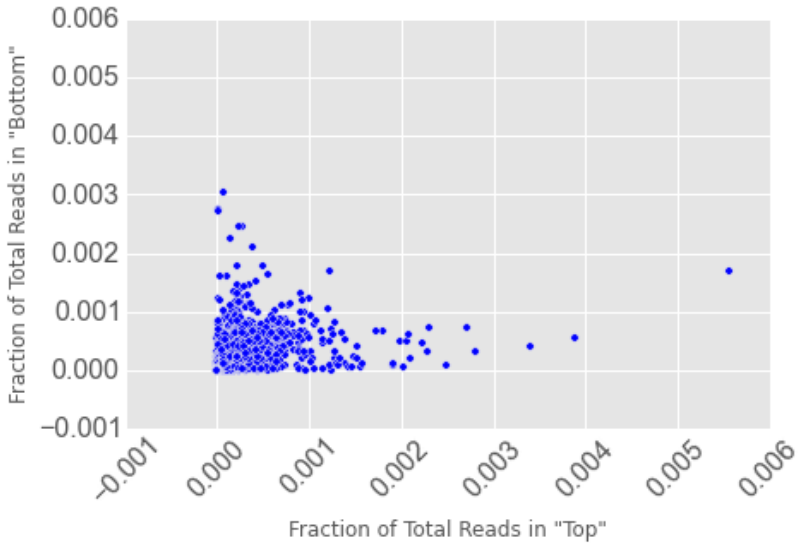

Figure 3.4. Variants enriched for high fluorescence are different from those enriched for low fluorescence. This demonstrating that fitness is not the major contributor to enrichment results. Values plotted are $\log_2$ ratios of barcode read counts for high and low fluorescence bins versus control (normalized).

Variants that produce more betaxanthin would be enriched in the Top sort and de-enriched in the Bottom sort. To fairly compare variants, the enrichment and de-enrichment need to be expressed as a single value. Previous work has established that a good metric to use is the "$\log_2$ ratio" of the read counts (Wetmore et al., 2015). To calculate these values, for each variant, I took the ratio of the read count in "Top" divided by the read count in "Bottom" and applied $\log_2$ to those values. I then normalized these measurements by subtracting the median $\log_2$ ratio to account for the fact that there may have been a difference in the number of total reads of the two samples being compared, which would artificially inflate or deflate these values(Wetmore et al., 2015). In this dissertation, I'll refer to the $\log_2$ ratio

values for the read counts in the “Top” versus “Bottom” samples as “enrichment scores”. Variants that produce more betaxanthin would have positive and large enrichment scores. Variants that have no change in betaxanthin compared to wild type will have an enrichment score near zero. Variants that have decreased betaxanthin production should have negative enrichment scores.

Figure 3.5 shows the $\log_2$ ratios for “Top” versus “Bottom” read counts. Each point represents one variant containing a single gene deletion. The entire sort-seq experiment was repeated twice and the enrichment scores in the two replicates have a Pearson correlation of 0.7. This was a promising finding, since it suggested that the sort-seq experiment could be performed consistently and the enrichment scores were unlikely to be explained by random variability in the cultures or measurement noise.

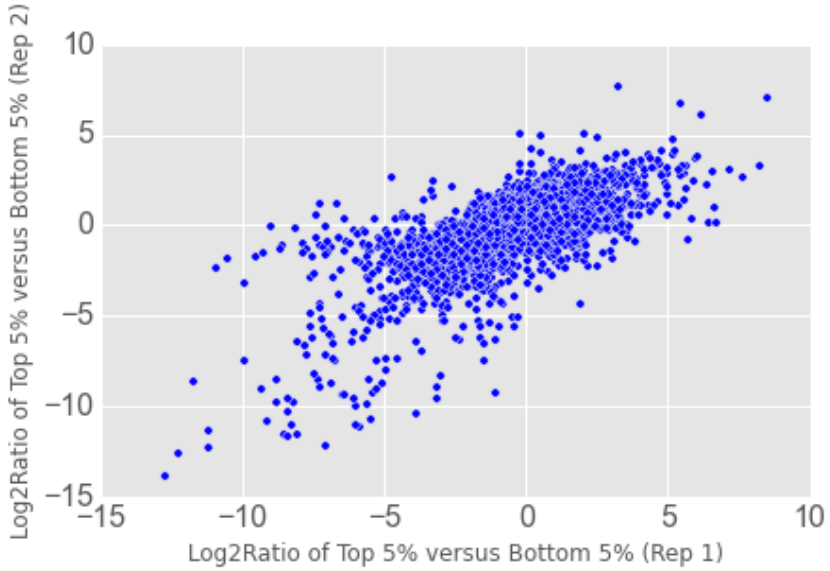

Figure 3.5 Two replicates of the sort-seq experiment give similar results. Blue circles represent individual ORFs. Values on the axes represent the $\log_2$ ratios of top versus bottom sorting bins in two different replicates. Points at the upper right represent ORFs that are highly enriched in the top sorted bin in both replicates of this experiment.

The genes with high enrichment scores in these two replicate experiments are presented in Tables 3.1 and 3.2. Most of the variants in the library had two barcode sequences associated with them, the UPTAG and the DNTAG (Brachmann et al., 1998). In these tables the tag with the higher enrichment score was used to determine the rank for the variant. In addition to measuring the enrichment score, I collected data before and after 24 hours of growth

without sorting to assess the fitness of the variants. These “fitness scores” were calculated as the $\log_2$ ratio of the after-growth counts versus the pre-growth counts. The top 20 variants in each replicate are shown here, though the fitness and enrichment scores were calculated for all members of the library.

Table 3.1 Variants with high enrichment scores in replicate 1

Deleted Gene	Enrichment Score	Fitness Score
PDR8	8.5	-0.08
COA1	8.26	1.75
LYS12	7.66	-0.29
YAF9	6.72	-1.55
CDC10	6.66	-11.55
YBR209W	6.57	-0.03
SWC3	6.34	-0.46
ZRT1	6.16	-1.03
BOP2	6.01	0.05
LYS2	5.93	0.25
ACO2	5.84	-1.98
HXK2	5.67	0.66
PTK2	5.66	0.03
TIR3	5.65	-0.72
YBR225W	5.59	0.35
MTC4	5.59	0.85
PCL6	5.48	0.24
LDB18	5.47	-2.44
YOR008C-A	5.46	-1.13
LYS1	5.37	0.84

Table 3.2 Variants with high enrichment scores in replicate 2

Deleted Gene	Enrichment Score	Fitness Score
CTS1	7.64	-2.59
PDR8	7.06	0.56
LDB18	6.78	-2.22
ZRT1	6.14	-1.31
MTH1	5.11	0.37
ACL4	5.01	-0.25
TRP3	4.8	-0.35
DLT1	4.71	0.56
YOR1	4.26	0.17
ERG5	4.19	-0.33
MTC2	4.15	-1.52
YPR050C	4.1	-1.77
MTC6	4.1	-0.59
EGD2	4.06	-2.09
MGS1	4.05	-0.09
QDR2	4.02	-1.31
ISW1	3.94	-0.22
TIR3	3.91	-0.56
RPS10B	3.9	-2.68
LYS12	3.84	0.2

While the top 20 variant lists for these two replicates show very different results, several of the gene deletions are common to both lists. For example, PDR8 appears at the top of both lists, matching the results found in the threshold sorting experiment (section 3.2). QDR2 is found at rank position 95 in the first replicate (not shown in Table 3.1) and at position 16 in the second

replicate. This result was surprising since we know that the deletion of QDR2 should increase fluorescence from the results of the threshold sorting experiment (section 3.2).

Gene deletions that are common to the top 20 variants in both replicates are PDR8, LDB18, LYS12, ZRT1, and TIR3. To generate a single list of hits for validation, I wanted to use the data from both replicates. However, in each replicate, data is missing for some of the variants that had high enrichment scores in the other replicate. Furthermore, the two replicates had markedly different values for enrichment scores. The phenotypes in replicate 1 were much more pronounced. That is, the values for the enrichment scores were more extreme at the high end for replicate 1 compared with replicate 2. Using the average enrichment scores between the two replicates would over-weight the first replicate. Thus, I chose to calculate the “mean rank” for each variant, which is the average of the rank position of the variant in the two experiments. In the cases where one replicate did not produce any data for a given gene (read counts were too low or of low quality), the rank was assigned as N/A and the average rank was assigned as the rank in the replicate with data.

Table 3.3 Variants with low mean rank across two replicates

Deleted Gene	Mean Rank	Rep 1 Rank	Rep 2 Rank	Rep 1 Enrich. Score	Rep 1 Enrich. Score
PDR8	1.5	1	2	8.50	7.05
ZRT1	7	10	4	6.16	6.14
YOR1	10	N/A	10	N/A	4.26
LDB18	13	23	3	5.413	6.78
MGS1	16	N/A	16	N/A	4.05
BOP2	17.5	11	24	6.01	3.77
DLT1	18	27	9	5.18	4.71
MTC2	19	26	12	5.26	4.14
QDR2	19	N/A	19	N/A	3.93
RPS10B	22	N/A	22	N/A	3.89
LYS12	23	N/A	23	N/A	3.83
COA1	23.5	2	45	8.26	3.27

MTC6	24.5	28	21	5.16	3.90
IFM1	25	N/A	25	N/A	3.71
PCL6	31	20	42	5.47	3.32
TIR3	32	17	47	5.65	3.24

The deletion collection is known to have a number of secondary mutations (Eason et al., 2004). It is possible that these secondary mutations cause the phenotype observed in a screen rather than the deletion that the barcode is associated with. Working with an undergraduate, Francesca-Zhoufan Li, I generated markerless, Cas9-mediated deletions of each of these genes in a wild type background strain that contained the betaxanthin biosensor genes. I then performed flow cytometry to validate that these deletions were, in fact, causing an increase in cell fluorescence. In Figure 3.6, fluorescence distributions for a subset of 8 of the deletions in Table 3.3 are shown in comparison to the no deletion control in black. These data show that the PDR8 and QDR2 deletion mutants still have an effect on betaxanthin fluorescence in the “clean” strain background. YOR1, PCL6, and TIR3 deletion mutants have a small increase in fluorescence compared with wild type. However, the other deletions have a small effect or a negative effect on fluorescence. This may indicate that secondary mutations or measurement noise may have accounted for the initial enrichment for these variants in the sort-seq experiment.

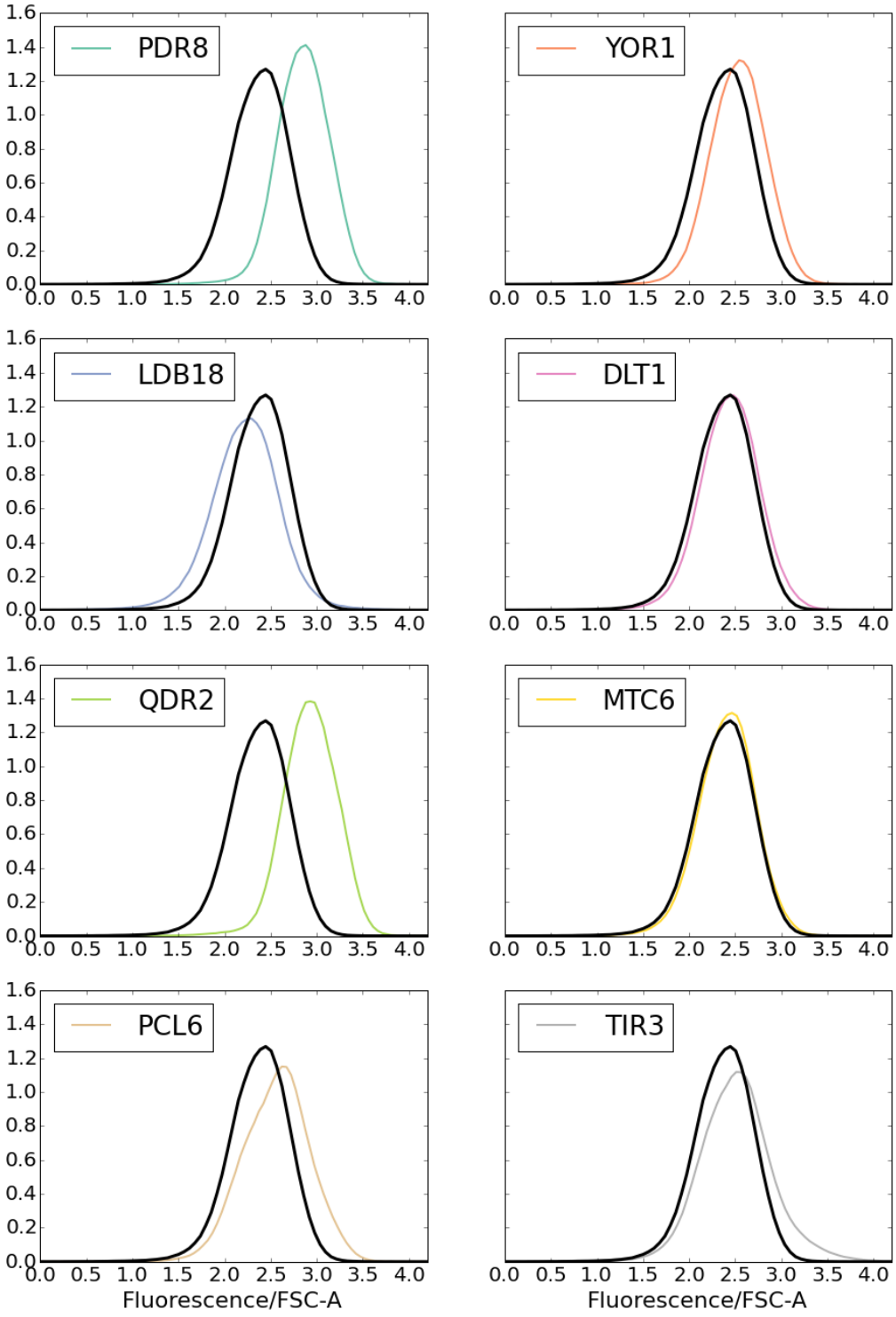

Figure 3.6 Fluorescence distributions for validation strains. Deletion strains in clean backgrounds were grown in the same media used in the sorting experiments. Fluorescence measurements were normalized for cell size using forward scatter area measurements. Black lines represent a

kernel density estimator for the size-normalized fluorescence distributions for a background strain containing no deletion. Each colored line represents a kernel density estimator for the size-normalized fluorescence distributions of each candidate deletion.

3.4 Validation of hits found in deletion collection

To confirm that the increase in fluorescence in the flow cytometry data above was due to increased L-DOPA production, undergraduate Francesca-Zhoufan Li and I generated strains containing CYP76AD1, feedback-resistant ARO4, and markerless deletions of some of the hits described in section 3.3. I grew liquid cultures of these strains and measured L-DOPA concentration in the culture supernatant using a colorimetric assay. The assay uses a solution called Arnow's reagent (Arnow, 1937). This reagent reacts with catechols, resulting in a pigment that absorbs light at 460nm. The assay could only be used to measure L-DOPA extracellularly. The strains produced similar or less L-DOPA as the wild type strain (WT) that had no deletions (Figure 3.7).

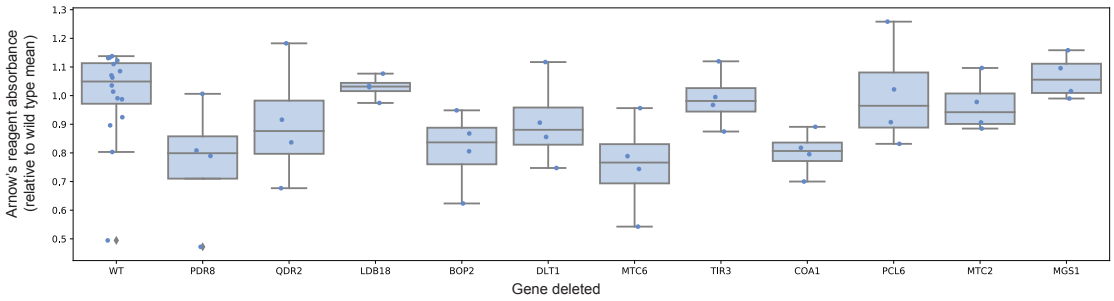

Figure 3.7 L-DOPA concentration in sort-seq hits as measured by Arnow's reagent assay. Absorbance of culture supernatant samples reacted with Arnow's reagent were measured at 460nm in a plate reader. WT represents a background strain containing CYP76AD1 to produce L-DOPA but lacking any gene deletions. Boxes represent the interquartile range (IQR; 25-75%). Whiskers represent the range of the data with the exception of outliers. Data points outside of the IQR +/- 1.5xIQR are considered outliers and are marked with diamonds.

While I was able to make a standard curve for this assay using commercially available L-DOPA, the linear portion of the standard curve and the absorbance values of the standard curve varied dramatically between replicates. Because of this, the data shown in Figure 3.7 are taken from OD-normalized absorbance data, without conversion to a concentration using a

standard curve. Furthermore, I observed that some of the samples formed precipitates (samples containing different strain supernatants). Because of these two issues, there remains some uncertainty about whether or not its measurements are accurate. The assay has only been described in a few prior publications and is not widely used.

L-DOPA elutes quickly in reverse-phase liquid chromatography due to its polarity and it is unstable in solution, so we were unable to measure L-DOPA concentration using LC/MS. However, an enzyme DOPA decarboxylase (DODC) from *Pseudomonas putida* can decarboxylate L-DOPA to produce the more stable product dopamine. Starting with a strain expressing heterologous CYP76AD1, feedback resistant ARO4, and DODC to produce dopamine, we created markerless deletions of most of the genes above. We were unable to generate deletions of MGS1 and COA1 in this strain before performing the dopamine assay. Using a similar protocol as with L-DOPA, I grew the strains and collected culture supernatants. We were able to quantify dopamine concentration using LC/MS (Figure 3.8). From the unpublished work of another member of the lab, we knew that dopamine is actively exported from yeast cells, so the signal detected in the supernatant should also be higher than for L-DOPA.

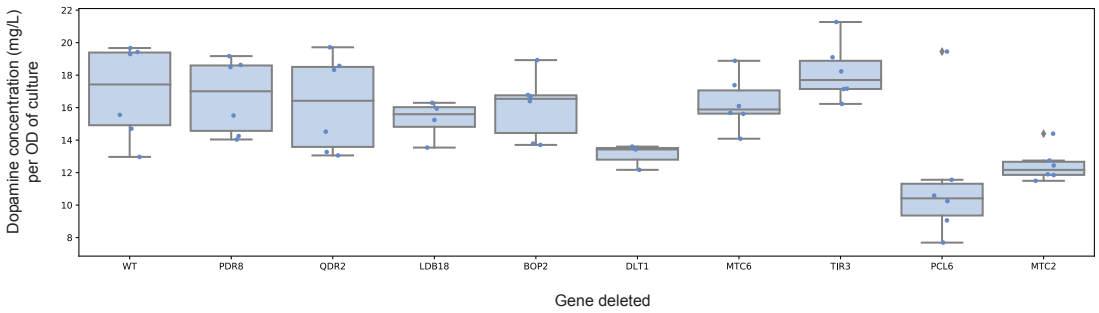

Figure 3.8 Dopamine concentration in sort-seq hits show no improvement over a strain with no deletions. Concentration of dopamine in culture supernatant samples was measured using LC/MS. WT represents a background strain containing CYP76AD1 and DODC but lacking any gene deletions. Boxes represent the interquartile range (IQR; 25-75%). Whiskers represent the range of the data with the exception of outliers. Data points outside of the IQR +/- 1.5xIQR are considered outliers and are marked with diamonds.

As shown in Figure 3.8, the dopamine production in the deletion mutants was within the error of the WT production rate. The differences between biological replicates were very significant. Wild type dopamine production, for

example, varied between approximately 13 and 19.5mg/L. Two of the deletions, PCL6 and MTC2 actually showed decreased dopamine production, rather than increased. Taken together, we decided that the hits emerging from the sort-seq screen are unlikely to be true positives.

3.5 Conclusions

The threshold sorting technique allowed us to identify that PDR8 and QDR2 deletions increase fluorescence in single-cell screening and do so by increasing intracellular relative to extracellular fluorescence in a liquid culture. The sort-seq results gave the same two hits in addition to a number of other candidates. However, only PDR8, QDR2, and to a lesser extent YOR1 actually increased fluorescence when deleted in a “clean” background strain. This suggests that the sort-seq methodology is highly noisy and enriches for many false positives. Still, results from the threshold sorting technique were more reliable than those in the sort-seq technique, so we used threshold sorting only in subsequent screens.

The exact source of this result is unknown. One hypothesis is that certain variants in the library were double transformants, meaning that they integrated multiple. If this is the case, these particular variants may cause produce more betaxanthin during the screening procedure due to their increased dosage of biosensor cassette genes. During validation, the mutations in these variants would be reproduced in a “clean” background strain that is known to contain a single copy of the biosensor gene cassette. If the gene dosage, rather than the mutation, was the cause of increased betaxanthin production, we would expect these mutants to fail upon individual validation as shown in Figures 3.7 and 3.8. This problem would persist in the case of threshold sorting as well, but gene dosage could be tested in enriched individuals since the live cells are recovered in this technique.

Another possibility to explain the absence of increased dopamine and L-DOPA production in sort-seq-enriched mutants could be that each mutation affected an aspect of betaxanthin production, compartmentalization, or degradation that is not relevant for L-DOPA or dopamine. For example, betaxanthin is primarily found in the vacuole (DeLoache et al., 2015). If a mutant in the library affects vacuole size or morphology, it will appear to change betaxanthin levels without affecting pathway flux more broadly. These types of false positives should affect both sorting techniques.

4 Building a transposon-based, barcoded, insertional mutagenesis library and screening for mutants with increased bioproduction of betaxanthin

In Chapter 3, I showed that deleting QDR2 increased the retention of betaxanthin inside the cell. We reasoned that future screens should be performed in a strain with QDR2 deleted (yJS1159) in order to maximize the retention of the fluorescence signal we use for sorting and reduce the transfer of betaxanthin between strains in our libraries. I attempted to perform a library-scale transformation to delete the QDR2 gene in the existing library that was used in Chapter 3 (deletion collection with betaxanthin biosensor genes). However, the transformation efficiency was too low to cover the entire library. Furthermore, we suspected that we would continue on multiple rounds of library generation and screening in the future. We decided to create an *in vitro* transposon-based disruption library, which would serve as a tool for generating deletion libraries in new background strains.

4.1 Introduction to transposon-based insertional mutagenesis and previous studies

The yeast deletion collection, a set of *S. cerevisiae* strains in which a single ORF has been deleted, has been a major resource for studying genomic deletions that confer desired phenotypes (Brachmann et al., 1998; Giaever & Nislow, 2014). The process for building the deletion collection required the participation of multiple labs over a number of years, so repeating the production of this library in new strain backgrounds is prohibitively challenging. This precludes researchers from using the deletion collection for iterative screening: the process of screening libraries that are produced using strains that have been updated with the results of previous rounds of screening. Thus, it is challenging to perform iterative rounds of screening for deletions.

One strategy for creating gene knockout libraries in new background strains is to use transposon-based disruption (Oh et al., 2010; Ross-Macdonald, Sheehan, Friddle, Roeder, & Snyder, 1999). To perform multiple iterations of library generation with transposons, each round would require the researcher to generate a new strain library and characterize the insertion sites of the transposons. Furthermore, these insertion sites will be different in each library, making cross-strain comparisons challenging. For this dissertation, in collaboration with Ryan Protzko, I created a DNA-barcoded, *in vitro* preparation of a transposon-disrupted genomic library. Since the

transposons were barcoded before they integrated into the genome, we only needed to characterize the insertion positions one time. After that characterization, we could use the same library in any *S. cerevisiae* strain and simply Sanger sequence the DNA barcode to identify mutants.

4.2 Methodology for building transposon-based, barcoded, insertional mutagenesis library

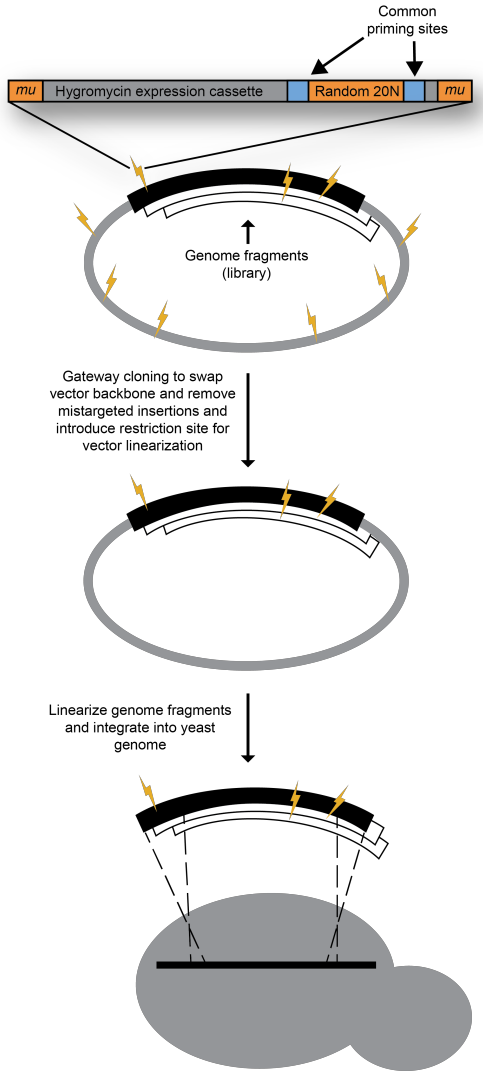

Fig 4.1 Schema for building *in vitro* transposon-based disruption library. Transposon insertion events are represented by yellow lightning bolts. Genomic fragments are represented by black rectangles and white rectangles outlined in black. Each transposon insertion event moves the

sequence of the mu transposon (top of figure) into the genome fragment plasmids. The process for generating the disruption library is described in detail in the text of this section.

We developed a process, summarized in Figure 4.1, to create and characterize insertions in yeast genomes *in vitro* before transferring these to live yeast cells. To build the *in vitro* transposon-based disruption library, we obtained a library of yeast genome fragments from the tiling library described in Jones *et al*. (Jones, Stalker, Humphray, Dunham, & Prelich, 2008a), which we purchased in a pooled format (Dharmacon, Lafayette, CO). This library comprised 1500 unique plasmids containing genome fragments from the *S. cerevisiae* genome of 1-20kb. In the plasmids, the genome fragments were flanked by Gateway cloning “attL” sites. The backbone of the plasmid contained a LEU2 selectable marker expression cassette for selection in yeast, a KanR expression cassette for selection in *E. coli* on kanamycin, a 2 micron yeast origin of replication, and a pUC origin of replication with a ROP (Repressor of Primer) site which controls copy number of the plasmid.

We customized a mu transposon encoded in a plasmid by cloning a trimR *E. coli* selectable marker and HygR *S. cerevisiae* selectable marker in between the two mu repeat sites within a ColE1 plasmid. Mu repeat sites were flanked by BamHI and HindIII restriction sites. We created trackable barcodes by cloning a 20bp barcode flanked by conserved priming sites directly before a mu repeat site. A library of random barcodes was generated by PCR-amplifying this plasmid using primers with 20bp of degenerate sequence and flanking SapI type IIs restriction sites. This PCR product was then circularized in a monomolecular SapI digest reaction. All cloning was performed using the golden gate method with enzymes BsaI, BsmBI/Esp3I, and SapI.

Transposons were excised from the transposon plasmid using a BamHI and HindIII digest. Transposons were integrated into the genomic fragment library using mu transposase in an *in vitro* reaction, according to manufacturer protocol (ThermoFisher, Waltham, MA). To remove transposon insertions that occur in the plasmid backbone from our library, we used a gateway cloning reaction to transfer the disrupted genomic fragments to a plasmid backbone with a different selectable marker (SpecR).

This library, named pJS1362L, was characterized by the RB-TnSEQ method to associate transposon barcodes to the insertion site of the transposon in the genome. In the sequencing data, we still found 45% of the barcodes associated with plasmid backbone DNA from the original genome fragment library. The remaining insertions were distributed equally among the

16 chromosomes (Figure 4.2A). Among the insertions (unique barcodes) that mapped to the original genome library plasmid backbone, a disproportionate number occurred in the ROP region of the plasmid backbone (Figure 4.2B). We reasoned that this was because disruptions in ROP break its regulation on the plasmid copy number, causing these plasmids to replicate at a faster rate than those with disruptions in other parts of the plasmid sequence. Moving forward, we ignored any barcodes that did not map onto an insertion site in the yeast genome.

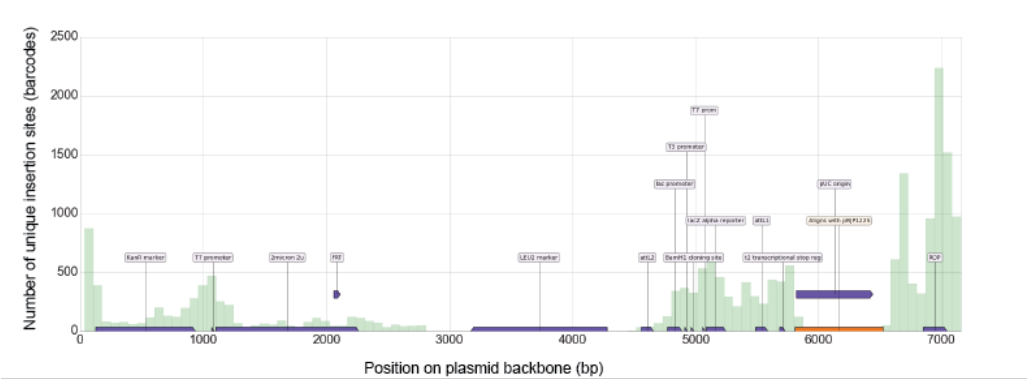

Fig 4.2 Transposon insertion sites in backbone of plasmid vector. Distribution of transposon insertion sites in the genomic fragment plasmid library from Jones *et al* (Jones, Stalker, Humphray, West, Cox, Rogers, et al., 2008b). A large fraction of the plasmids sequenced had insertions in the ROP region of the plasmid backbone.

In the new plasmid backbone, flanking the genome fragments, were SceI homing endonuclease restriction sites, which are 18bp long and do not occur in the yeast genome. We used these sites to excise the transposed genome fragments and integrate them into yeast genomes to disrupt native genes. Yeast cells containing transpositions were selected for using hygromycin.

4.3 Characterization of library

In our single library preparation, we sequenced the transposon insertion sites using RB-TnSEQ (Wetmore et al., 2015). For each gene in the genome, we considered a transposon insertion event to be a disruption in a transcribed feature (either an ORF or RNA) if it fell within the first 75% of the feature or within 500bp upstream of the feature. These heuristics were used to account for the promoter region preceding the gene and the fact that some genes can still be functional without their 3' end. In this preparation, 58%

(4202/7198) of transcribed features had at least one transposition event satisfying these heuristics.

4.4 Second round of screening for increased betaxanthin production

We linearized and integrated this transposon-mediated disruption library into a strain containing the betaxanthin biosensor genes and *Δqdr2* (yJS1159) by native homologous recombination. We used FACS to identify mutants within this library with higher fluorescence and identified mutants containing transposon disruptions of the transporter gene YOR1. The sort-seq results in our first round of mutagenesis and screening (Section 3.3) corroborate this result, as YOR1 was one of the strains with a high enrichment score. We verified the phenotype of this YOR1 disruption by building a markerless, Cas9-mediated deletion of YOR1 in our background strain, with and without the QDR2 deletion. After measuring intracellular versus extracellular betaxanthin in liquid cultures of this strain, we hypothesized that deletion of YOR1 further decreases efflux of betaxanthin from the intracellular environment (Figure 4.2). Furthermore, we verified that combining the markerless deletions of QDR2 and YOR1 in our background strain had an additive effect in compartmentalizing BX. In a *Δqdr2 Δyor1* strain, generated in a clean background, approximately 78% of the betaxanthin was retained intracellularly (Figure 4.2).

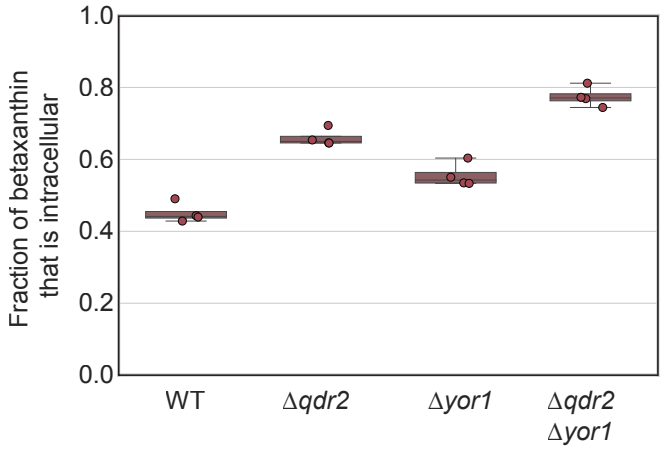

Figure 4.2 Iterative screening using a transposon-mediated disruption library in the *Δqdr2* strain and decreased CYP expression identified another betaxanthin transporter: YOR1. A transposon-disruption library in a Δqdr2 background enabled identification of a second betaxanthin export mechanism: YOR1. *Δyor1* strain retained more

betaxanthin than wild type. The *Δyor1Δqdr2* double mutant strain retained 77.4% +/- 2.8% of betaxanthin produced intracellularly, compared with 45.1% +/- 2.7% in the wild type strain. All four strains contain the same betaxanthin biosensor gene cassette, but expression of CYP76AD1 was driven by pRPL18B, which is approximately 10-fold weaker than the promoter, pTDH3, used to drive expression of CYP76AD1 in the strains shown in Figure 2. Boxes represent the interquartile range (IQR; 25-75%). Whiskers represent the range of the data with the exception of outliers. Data points outside of the IQR +/- 1.5xIQR are considered outliers.

4.5 Third round of screening for increased betaxanthin production

With a *Δqdr2 Δyor1* strain background, we hypothesized that we would be able to uncover beneficial disruption mutants that were difficult to isolate in previous rounds due to betaxanthin export. In this third round of screening, we integrated the transposon-mediated gene disruption library into a *Δqdr2 Δyor1* biosensor strain (yJS1256) and discovered that the disruption of heme oxygenase HMX1 increased betaxanthin production by approximately 5.2% +/- 0.9% (SEM; Figure 4.3). To verify this disruption, we generated a markerless Cas9-mediated deletion of HMX1 in our background strain. Validating this result by measuring L-DOPA is precluded by L-DOPA's instability during culture extraction and mass spectroscopy. However, its decarboxylated product, dopamine, is far more stable, can be more easily extracted from spent media. We added the gene DOPA decarboxylase (DODC) to measure dopamine production in these strains and found a 28.4% +/- 10.5% (SEM) increase in dopamine production with *Δhmx1*.

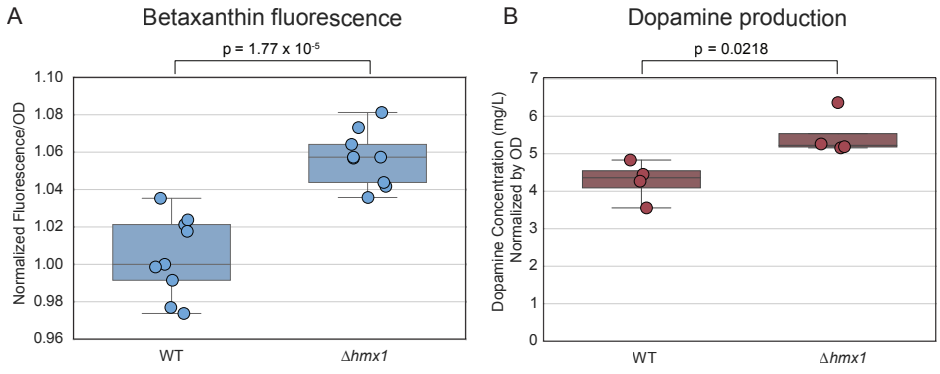

Figure 4.3 Third iteration of screening identifies *Δhmx1*. Screening of a transposon-disruption library in a Δqdr2 Δyor1 background led to identification of Δhmx1, a disruption of heme oxygenase, that increased both betaxanthin and dopamine titers in their respective production strains. A) The *Δhmx1* strain shows higher total fluorescence than WT in liquid culture. B) The *Δhmx1* strain produces 28.4% +/- 10.5% (s.e.m) more dopamine than WT after a 14h growth period. Boxes represent the interquartile range (IQR; 25-75%). Whiskers represent the range of the data with the exception of outliers. Data points outside of the IQR +/- 1.5xIQR are considered outliers. Significance was determined using t-tests.

4.6 Conclusions

The transposon-based disruption library that Ryan Protzko and I created gave us the unique ability to screen for deletion mutants with increased fluorescence through multiple rounds of mutagenesis and screening, updating the background strain each time to include the results from previous rounds. Through this iterative screening we were able to improve the screen itself (by compartmentalizing the signal molecule, betaxanthin). The improved screen enabled us to identify that deleting HMX1 increases betaxanthin fluorescence and also increases dopamine production.

5 Investigating the role of HMX1 in bioproduction via cytochrome P450 enzymes

5.1 Deletion of HMX1 increases intracellular heme concentrations

HMX1 encodes heme oxygenase, which is involved in the degradation of heme and regulation of several antioxidant-defense enzymes. We found that deletion of HMX1 increases heme concentration in wild type yeast cells 2.2 +/- 0.4-fold (SEM) (Figure 5.1). As a comparison and positive control, we also measured the heme concentration in a strain overexpressing heme biosynthesis genes HEM2, HEM3, and HEM12, a known strategy for increasing heme(Michener et al., 2012). The overexpression of heme biosynthesis genes has a much larger effect on heme concentrations than the deletion of HMX1. In Figure 5.1, heme measurements were made using an oxalic acid-based assay described in the methods section. The assay uses the intrinsic fluorescence properties of porphyrins and accounts for all heme in the cell, both protein-bound and free.

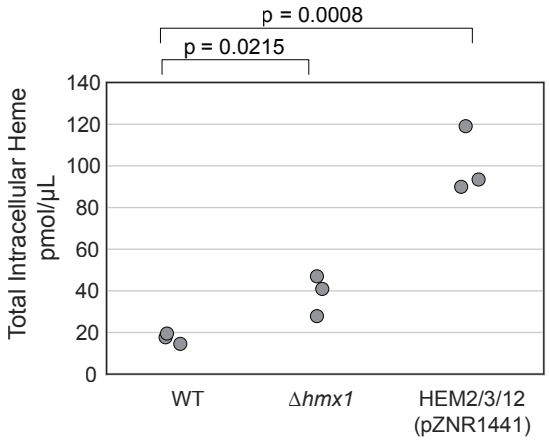

Figure 5.1. *Δhmx1* increases intracellular heme concentration. The deletion of heme oxygenase HMX1 increases the total intracellular heme concentration by 2.2 +/- 0.4-fold (s.e.m). A previously published strategy of overexpressing heme biosynthesis genes HEM2, HEM3, and HEM12 shows a larger increase in intracellular heme concentration of 5.9 +/- 0.7-fold (s.e.m). Significance was determined using t-tests.

5.2 Deletion of HMX1 increases activity of one other CYP-based pathway

We hypothesized that the deletion of HMX1 might be a generalizable strategy for improving the activity of heme-limited P450s. To test this, we integrated CYP102A1 (BM3) into a *Δhmx1* strain. BM3 catalyzes the conversion of indole to indoxyl, which then spontaneously dimerizes to form colored indigo(Jung et al., 2011; Michener et al., 2012). Indigo production increased by 22% +/- 7% (SEM) in the *Δhmx1* strain compared with wild type, similar to the improvement observed for CYP76AD1 F309L W13L (Figure 5.2). Previous work has shown that overexpressing HEM2, HEM3, and HEM12 increases indigo production from indole in a strain expressing BM3 (Michener et al., 2012). These two observations suggest that deletion of HMX1 may represent a common strategy for improving P450 function in yeast.

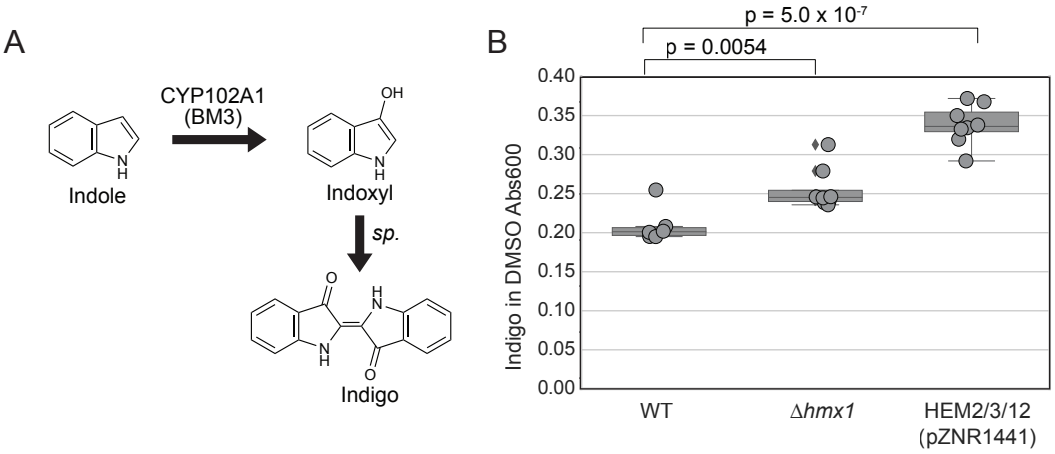

Figure 5.2 *Δhmx1* also improved activity for a second P450 enzyme. The deletion of heme oxygenase HMX1 improves flux through another P450 enzyme, CYP102A1 (BM3). A) CYP102A1 oxidizes fed indole to indoxyl. Indoxyl spontaneously dimerizes to form indigo, which is dark blue and precipitates out of solution. B) Strains containing CYP102A1 (BM3) produce indigo when fed indole. The *Δhmx1* strain produces approximately 22% +/- 7% (s.e.m) more indigo than the WT strain. Six samples were taken for WT condition and 8 samples were taken for the *Δhmx1* condition. Boxes represent the interquartile range (IQR; 25-75%). Whiskers represent the range of the data with the exception of outliers. Data points outside of the IQR +/- 1.5xIQR are considered outliers and identified with diamonds next to the marker. Significance was determined using a t-test.

5.3 Conclusions

Using the strain with increased betaxanthin compartmentalization (Δqdr2Δyor1), a third round of screening led to the discovery that disruption or deletion of heme oxygenase in *S. cerevisiae* leads to increased L-DOPA production. The deletion of heme oxygenase, which is reported to be involved in heme degradation, increases heme levels 2.2 +/- 0.4-fold (s.e.m). Furthermore, addressing heme limitation through deletion of HMX1 or overexpression of heme biosynthesis genes also increased activity of a different P450 enzyme, BM3, which is used in the production of indigo from indole. Taken together, these data suggest that the L-DOPA and indigo pathways tested here may be heme-limited, corroborating previous studies of P450 activity limitations (Krainer et al., 2015; Michener et al., 2012).

6 Building a barcoded yeast ORF overexpression collection and screening for betaxanthin production

6.1 Introduction and review of previous overexpression library studies

Screening deletions has a clear purpose – if the phenotype is absent in a particular deletion mutant, then that gene is likely responsible for or involved in the phenotype. Overexpressing genes and screening for a phenotype would uncover different types of useful genes. If a more extreme phenotype emerges in an overexpression mutant, then the gene is likely related to the phenotype and in some sense limiting for that phenotype. In the example of the betaxanthin screen presented here, overexpression mutants that increase betaxanthin production could have increased production of a limiting substrate or cofactor in the pathway.

In yeast, several groups have generated gene overexpression libraries (Gelperin, 2005; Ghaemmaghami et al., 2003; Huh et al., 2003; Jones, Stalker, Humphray, Dunham, & Prelich, 2008a). These studies used a few different strategies for library construction. Jones *et al.* cloned genome fragments at the 5-20kb size range into a plasmid vector (Jones, Stalker, Humphray, Dunham, & Prelich, 2008a). Gelperin *et al*, Huh *et al*, and Ghaemmaghami *et al*, all used PCR to amplify genes and insert them into vectors with different modifications: a HIS tag (Ghaemmaghami et al., 2003), a GFP fusion (Huh et al., 2003), a barcode(Ho et al., 2009), and an inducible promoter (Gelperin, 2005). A different strategy called TRMR has been used to make overexpression mutants in *E. coli*. In this strategy, authors used synthetic oligos to insert strong promoter sequences upstream of genes in the *E. coli* genome (Warner et al., 2010). To our knowledge, none of these tools have been used to increase the production of a valuable metabolic product. In fact, only the TRMR study studied a phenotype relevant for metabolic engineering: tolerance to lignocellulosic hydrolysate.

6.2 Building, sequencing, and characterization of overexpression library

In this study, we wanted to build a barcoded ORF library. Unfortunately, the existing barcoded library, the MoBY-ORF library from Ho *et al*, had genes expressed from their native promoters. We wanted to express genes from a strong, constitutive promoter in order to achieve high effect sizes and overcome any transcriptional regulation. The overexpression library built in Warner *et al* comprises ORFs flanked by attB sites. The library is called the Movable ORF Library (MORF) because gateway cloning reactions can be

used to move ORFs into commercially available or custom-built vectors (Gelperin, 2005). We choose to use the MORF library to transfer ORFs into a custom-built vector containing a strong, constitutive promoter, pTDH3, and a 20bp DNA barcode flanked by common priming sites. We also included homologies to genomic DNA to integrate the libraries, rather than expressing the genes from a plasmid. The construction process and final vector map is shown in Figure 6.1.

A key insight for cloning this plasmid library was that the barcode sites needed to be within 50bp of the inserted ORFs in order for next gen sequencing (using the RB-TnSEQ method) to be able to able to capture both in a single read. When both can be sequenced in one read, the barcode can be associated with the ORF that is present on the same plasmid. However, the barcode must be placed in the destination vector since adding it requires PCR which would be heavily biased when amplifying a set of plasmids with variable length genes in them. Therefore, the scars from the gateway cloning reaction required to move the ORFs would be between the ORF and its barcode.

In the gateway cloning system used with the MORF library, there are 4 “gateway sites”, which are also scars of their recombination reactions: attB, attP, attL, and attR. The attB site is the only one short enough to allow the barcode and ORF to be captured in one sequencing read. attB sites are formed by the reaction of attL and attR sites. However, the MORF library that is commercially available uses attB sites flanking the ORFs. I moved the ORFs into attL sites by a reaction of the MORF library with an interim destination vector containing attP sites. When attB and attP sites recombine, they form attL sites. Thus, the destination vector in Figure 6.1 contains attR sites. When these recombine with the attL ORF library, attB sites are formed to make the final library.

This library was characterized by RB-TnSEQ and used in homologous recombination reactions to transform yeast with these ORF libraries. Approximately 3560 ORFs (63.6% of the library) were mapped to a unique barcode that could be used to track them.

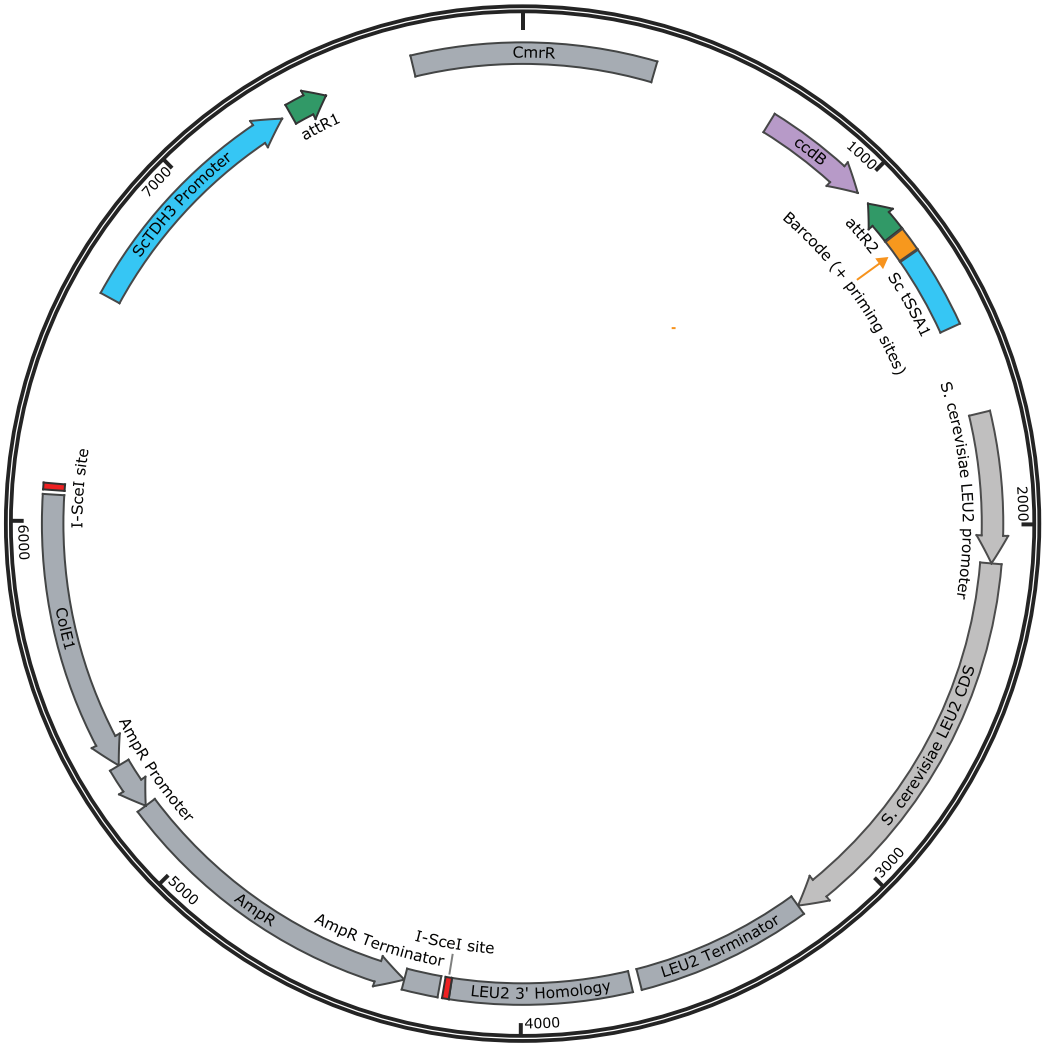

Figure 6.1. Overexpression library ORF entry vector pJS1426. In order to use the existing movable ORF (MORF) collection, I cloned pJS1426, which has a gateway cloning site to accept the gateway-compatible ORFs. The gateway reaction places the ORF between the TDH3 promoter and SSA1 terminator, shown in blue. The attR sites to be used for recombination are shown in green. The location of the barcode (orange) is between the attR and tSSA1 sites; it contains a stop codon to end translation before the terminator. The I-SceI sites, labeled in red, are 18-bp unique restriction sites and enable linearization of the plasmid.

6.3 Threshold sorting of overexpression library

The overexpression library was integrated into a *Δqdr2* background with the biosensor cassette (yJS1159). This experiment was performed before the complementary effect of YOR1 on compartmentalization was discovered. I sorted this population using the threshold sorting methodology (top 0.5%, then verify in liquid cultures) to screen this population. I picked the colonies that appeared brightest on an agar plate into liquid media and measured fluorescence in a plate reader. Figure 6.2 shows the fluorescence measured for each culture, labeled with the well number of the culture in the plate. The average fluorescence for the strain with an empty overexpression plasmid is shown as a horizontal black line. Figure 6.3 is the same data reproduced but highlighting the top eleven overexpression mutants. I sequenced the barcodes of these mutants and identified the genes. They are listed on the x-axis in Figure 6.3.

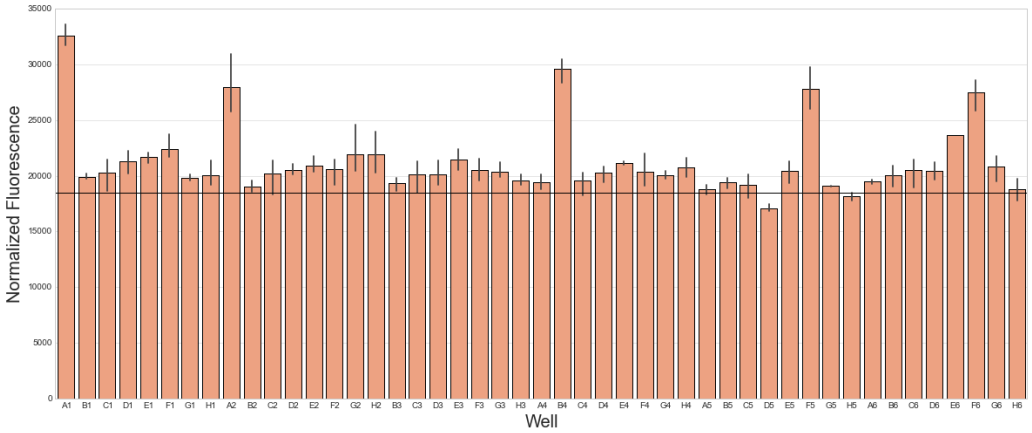

Figure 6.2. Liquid culture fluorescence of all enriched overexpression mutants. Cultures were grown from single colonies in synthetic dextrose media lacking uracil and tyrosine. The y-axis refers to the fluorescence of the culture at excitation/emission 485/525, normalized to the optical density as a proxy for cell count.

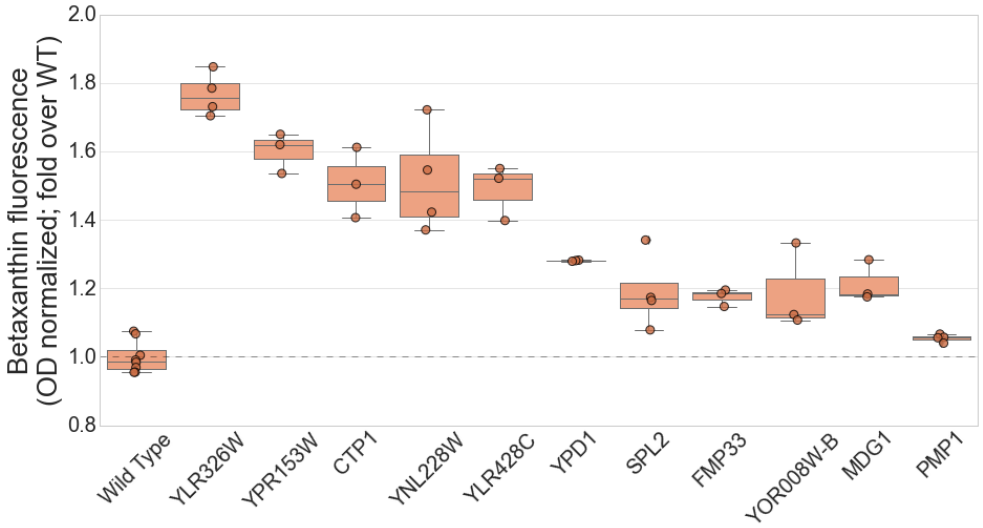

Figure 6.3 Liquid culture fluorescence of highest fluorescence mutants identified by gene overexpression. Highest fluorescence strains from Figure 6.2. Gene identities were determined by BarSEQ reactions amplifying the barcodes near overexpressed ORF sites.

For the eleven overexpression mutants showed in Figure 6.3, I researched past literature to understand the known roles of these genes. For all of them, it was unclear what mechanism would be causing the increased betaxanthin fluorescence phenotype.

6.4 Validation of hits from threshold sorting of overexpression library

Since the strains in Figures 6.2 and 6.3 were derived from a pooled library, we needed to recreate them in clean background strains to verify that the overexpressed gene caused increased betaxanthin production, rather than a background mutation. Francesca-Zhoufan Li and I built a strain with CYP76AD1 and feedback-resistant ARO4 to produce L-DOPA. The top five candidates shown in Figure 6.3 were integrated into the LEU3 locus of this strain as was done in the library.

By this time, I had discovered a new colorimetric assay for L-DOPA using the chemical resourcinol (J. Zhao et al., 2017). Resourcinol fluorescence is proportional to the concentration of L-DOPA present in a sample, but it is insensitive to tyrosine or dopamine. Figure 6.4 shows that the L-DOPA concentrations in the five overexpression strains were similar to the concentration in wild type, even though their fluorescent betaxanthin production was much higher in the context of strains derived from the library.

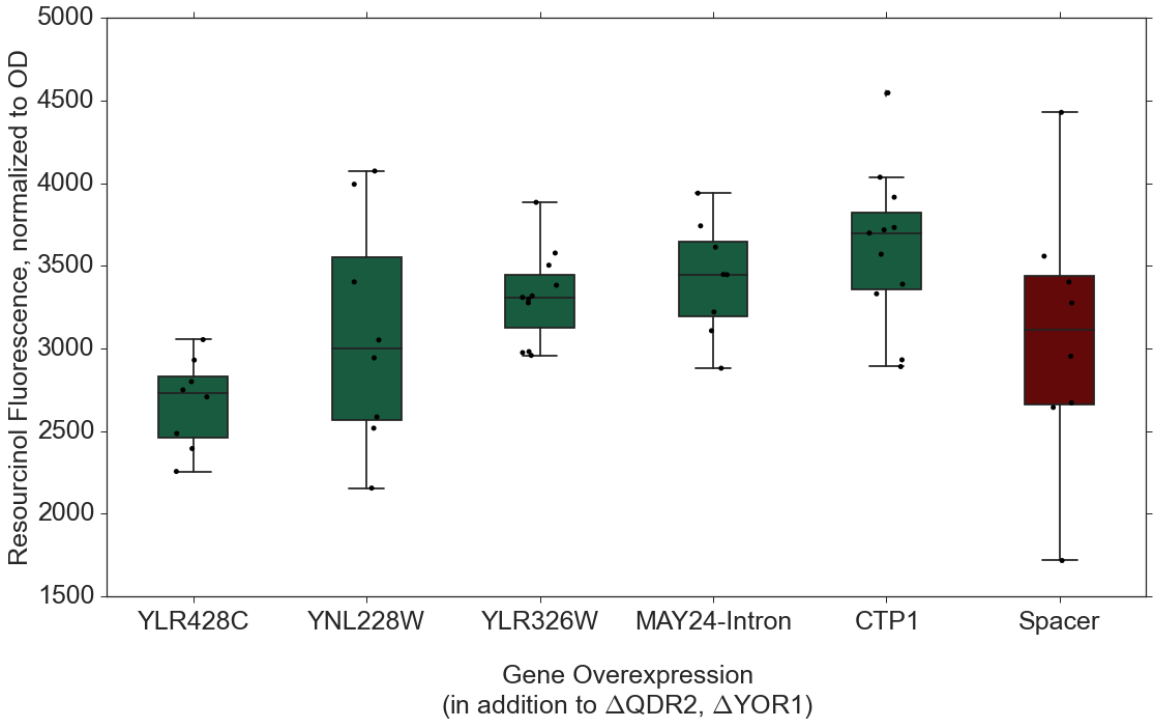

Figure 6.4 Overexpressed gene hit candidates did not increase L-DOPA production when integrated into a clean background strain. Strains with overexpressed genes in non-library (clean) background strains were grown as in Section 6.3 and measured for L-DOPA concentration using the resourcinol assay. The y-axis represents the fluorescence of the product of resourcinol reacting with L-DOPA, normalized by the optical density of the culture as a proxy for cell count.

6.5 Conclusions

We were successfully able to generate an overexpression library in yeast that contains 1) homology arms for genomic integration, 2) a high expressing promoter, and 3) a barcode for tracking the ORF. While the library is incomplete (64.6% of ORFs are represented), the methodology can be used to create a new library with any of the plasmid features customized.

Screening the library produced four promising hit candidates, but upon further investigation, none were found to strongly increase L-DOPA production in yeast.

7 Discussion

7.1 Discussion of results

The iterative screening strategy employed in this study was enabled by a barcoded, *in vitro* transposon-based disruption library. Construction and characterization of an *in vitro* library allowed us to perform multiple rounds of screening with updated background strains in each round without needing to rebuild the library for each round.

Employing this library, we used iterative host mutagenesis and FACS screening to first identify mutants with improved single-cell screening (*Δqdr2*, *Δyor1*) and, subsequently, a mutant with improved P450 performance (*Δhmx1*). In our initial rounds of screening we discovered that fluorescent product compartmentalization is critical for the identification of mutations that increase production because it lowers coefficient of variation for the single cell measurements using this biosensor. We discovered two efflux transporters responsible for exporting approximately 72% +/- 6% (s.e.m) of the betaxanthin produced by the cells. Deletion of these two transporters increased compartmentalization of the fluorescent signal within single cells and enabled screening with a higher signal to noise ratio.

The challenge with library screening and selection in metabolic engineering is that it is possible that none of the library members have a strong impact on the phenotype of interest. From this study, as well as unpublished work of our collaborators, it appears that the libraries of single gene deletion and overexpression mutants do not contain any mutants that have a large effect on flux to L-DOPA production. With the exception of Δ*hmx1*, no library members queried in this study showed measurable improvement in pathway product production, despite several individual members showing increased fluorescence in the context of a pool of variants. We believe this indicates two key findings: 1) there are no single deletion or overexpression mutants of yeast that have a large effect on flux to tyrosine and functionality of the P450 gene CYP76AD1 and 2) during FACS enrichment, betaxanthin measurements are likely to produce false positives. It's unclear if the false positives were the result of noisy betaxanthin production or if the measurements in our FACS methodology were too noisy for discerning signal increases rather than on/off signals that FACS instruments are usually used for. However, the library generation techniques developed in this dissertation could be applied to a number of other metabolic pathways of interest.

7.2 Proposed next steps

The most promising direction for this work is to repeat screening for a different pathway that contains a fluorescent intermediate or a metabolite that can be converted into a fluorescent molecule in as few steps as possible. The localization and efflux of the fluorescent metabolite are not critical parameters since FACS screening can be used to identify mutants in which the metabolite is localized to the cytoplasm and able to give a strong enough signal for single-cell sorting.

Using the *in vitro* barcoded transposon-based disruption library, screening can be performed in multiple rounds; early rounds may enrich for hits that improve the screen and later rounds may enrich for candidate gene deletions that achieve a desired phenotype. Since the library has already been built and characterized, no further library generation would be required to perform this screen in a new pathway. However, as discussed in sections 4.2 and 4.3, the current library construction, pJS1362L is incomplete and contains a large number of backbone-only plasmids. We believe that both of these problems stem from the original gene fragment library used as a substrate: it was likely incomplete due to prior passaging in *E. coli* and its backbone contained features that, when disrupted, cause rampant plasmid amplification. Future applications of the *in vitro* barcoded transposon-disruption library would benefit greatly from rebuilding the library starting from a clean unbiased fosmid library of the yeast genome. Barcodes enable future screens or selections to quantify the effects of each mutant in parallel using only simple BarSEQ assays. Furthermore, analyzing read counts for each mutant enables the quantification of the degree to which each mutant displays the phenotype of interest, rather than a simple binary output.

Similarly, the overexpression library built and characterized in this study can be used for high-throughput screening or selection in yeast. Most existing overexpression libraries are not barcoded. This means that a screen or selection can only give information about a few individuals that are isolated and sequenced using Sanger methods. Furthermore, these results are binary – individuals are isolated or not, without a quantification of the degree to which they contain the phenotype or interest. Barcoding enables the parallel scoring of all mutants in the library for their enrichment in selective or screening conditions. The one existing barcoded overexpression library, MoBY, uses the native promoters of the overexpressed genes, so the addition of the library to the genome doubles the DNA copy number, but may not necessarily produce large increases in protein copy number (Ho et al., 2009). Gene regulation may mask the metabolic effects of these genes, so

we chose to generate a library with a strong, constitutive promoter controlling the overexpressed genes. We believe this library is more likely to produce a hit candidate in screens or selections because each library member is more likely to have increased protein production of the cloned gene and, therefore, produce a phenotype.

8 Broader context

8.1 Systems biology approaches can be used to improve engineered biosynthesis hosts

In metabolic pathways, heterologous enzymes and molecules are produced in a microbial host that is not evolutionarily adapted to their presence. These microbes have been studied to varying degrees, but even the most heavily studied hosts, *E. coli* and *S. cerevisiae*, cannot always be rationally engineered. We still lack a complete understanding of their metabolism and gene regulatory networks. The native biochemistry and gene regulation of the host may interact with heterologous pathways in unpredictable ways.

When rational engineering is insufficient, mutagenesis and screening/selection techniques can be used to identify unpredictable mutations or sets of mutations that improve pathway production. In the past few decades, random DNA mutagenesis at the level of individual base pairs (henceforth referred to as “point mutagenesis”) has been used extensively to identify mutants with improved productivity in both industrial microbes and crop plants. This was largely due to the availability of point mutagenesis technologies like ethyl methanesulfonate (EMS), X-rays, or UV irradiation (Foster, 1991). More recently, advances in genetic engineering and DNA synthesis, especially chip-based oligo synthesis, have enabled scientists to create libraries with a set of designed, targeted mutations (Bao et al., 2018; Cambray et al., 2018; Kosuri et al., 2013; LeProust et al., 2010; Mansell, Warner, & Gill, 2013; H. H. Wang et al., 2009). These systems biology techniques have enabled high throughput screening of gene dosage mutants. Mutant libraries can be created with gene deletion, overexpression, or reduction in expression.

There are several limitations to the use of “point mutagenesis” techniques. First, identifying the causal mutation among the many that are generated by these techniques is a challenge. Second, the mechanism by which a particular mutation produces the desired phenotype is not always easy to understand without further experimentation. When the mechanism of action

for a mutation is unknown, the insight cannot be used to produce improvements in other, unrelated metabolic pathways. Third and most critically, gain-of-function mutations are rare when using these methods. Gene dosage mutations, on the other hand, are easier to identify (for example, through barcode sequencing). They may be easier to understand and rationalize, as compared to point mutations and mutations in intergenic regions. New insights learned through these gene dosage mutants can be more easily applied to other pathways that share, for example, an enzyme class or a cofactor. Gene dosage mutations can include gain-of-function mutations through gene overexpression. Gene dosage mutations, versus point mutations, can be part of a systems biology approach to mutagenesis and selection since their effects can be directly reflected in metabolic models of the host organism. Lastly, libraries of gene dosage mutations can be used to discover phenotypes for many genes in the genome simultaneously producing data that can be used for constraining whole genome systems biology models. In summary, these types of mutant libraries enable systems biology to be leveraged for pathway insights and produce contributions to a systems biology description of the host organism.

In the specific case of engineering hosts for heterologous pathway production, I believe the systems biology approaches to mutagenesis and screening can be more valuable than point mutagenesis approaches. Toxicity or lack of productivity in a heterologous pathway are commonly caused by toxic intermediates, substrate limitations, interactions with host proteins, and improper trafficking (M. M. Zhang, Wang, Ang, & Zhao, 2016). With a few exceptions, many of these phenomena can be addressed by modifying the expression of proteins in the host genome. Substrate or cofactor limitations can be addressed by increasing production or decreasing degradation of these molecules. Pathway proteins or metabolites that get sequestered through interactions with host proteins can be restored by reducing or eliminating the offending protein from the host strain (if it is non-essential). Point mutagenesis can also break an unwanted interaction by mutating a residue required for the interaction. However, if the protein is essential, this mutation must not affect the protein's overall fold and any other functional domains. Such a mutant may not exist for the protein in question. Improper trafficking may be addressed or diagnosed by modifying the levels of trafficking and signal recognition proteins in the host. While a number of these issues could also be addressed through random mutagenesis, mutations that modify gene dosage can be used to address or at least diagnose all of these issues in a single experiment.

8.2 Pathway productivity can be improved through multiple on- and off-pathway mutations

Previous studies suggest that multiple mutations must be made in combination to optimize a host strain to produce a heterologous product (Gold et al., 2015; N. E. Lewis, Cho, Knight, & Palsson, 2009). Most studies that employ a systems biology approach to mutagenesis and screening use libraries in which mutants only harbor a single gene deletion or single gene overexpression (Giaever & Nislow, 2014). However, multiplexing of gene dosage mutations has been gaining interest. For example, one group demonstrated global transcription machinery engineering (GTME) in which the expression of many genes is altered simultaneously through the perturbation of transcription factor networks in the host genome (Alper & Stephanopoulos, 2007).

Screening mutant libraries in which each strain contains multiple mutations creates a problem at the stage of mutant identification. It is not always clear if one or multiple of the mutations were required for the phenotype. I believe that sequential, iterative mutagenesis can be used to account for this problem. Engineering strains with multiple mutations is challenging due to the lack of selectable markers, the evolution of resistance to transformation, and the laborious nature of library construction and characterization. Techniques that perform sequential mutagenesis without requiring new selectable markers or library construction may be powerful tools for finding multigenic mutants for metabolic engineering applications.

Iterative mutagenesis and screening with an *in vitro* transposon-based disruption technique can be used in any organism that can perform homologous recombination. The library would need to be created using genome fragments from this organism.

Other techniques have been proposed for generating deletion libraries, but none were able to do so multiple times in new background strains with requiring recharacterization. During the study described here, another method called CHAnGE was published that can be used to make genomic mutations for iterative mutagenesis and screening. The strategy uses plasmids containing sgRNA and donor DNA sequences to produce Cas9-mediated deletions at specified genomic positions (Bao et al., 2018). Both strategies have advantages and disadvantages but are roughly equal in financial cost and time required.

As high-throughput assay and sequencing technologies continue to become easier and cheaper to implement, I predict we will see an increase in the use of systems biology approaches for applications in synthetic biology

and metabolic engineering. These new approaches will allow researchers to apply their work beyond the metabolic pathway they are working on in any given project. They will allow the field to grow its base of knowledge about the properties of different organisms as metabolic engineering host strains in general.

9 Additional Materials and Methods

9.1 Strains

All strains were derived from *S. cerevisiae* BY4741. The biosensor cassette was comprised of three genes: 1) *Bv*CYP76AD1 F309L W13L, a cytochrome P450 gene derived from *Beta vulgaris* (beet) containing two mutations for enhanced activity that were identified in a previous publication (DeLoache et al., 2015); 2) *Mj*DOD, dopa dioxygenase derived from *Mirabilis japonica* (4-o'clock flower) that converts L-DOPA to betalamic acid, a precursor to fluorescent betaxanthin; and 3) *Sc*ARO4 K229L, DAHP synthase from the shikimate pathway containing a mutation that inactivates feedback inhibition by tyrosine (Hartmann et al., 2003). After adding the *Δqdr2* mutation following the first screen iteration, the betaxanthin signal was very high. Since we were able to identify improvements from *Δqdr2*, we could ensure our ability to detect further improvements by lowering the betaxanthin signal strength to the levels in the first screen iteration. Thus, strains and libraries built after the first screen iteration contained CYP76AD1 F309L W13L under the control of a weaker promoter than the one used for the first iteration of the screen, pRPL18B (weaker) versus pTDH3 (stronger).

9.2 Preparation of libraries for sorting

Library aliquots were thawed on ice for 30 minutes. The total aliquots were then transferred to 50mL of synthetic complete media containing 2% dextrose (SD) but lacking uracil and tyrosine. Cultures were grown at 30ºC with shaking for 6 hours. After growth, cultures had an optical density of 9.0-10.0. 8mL aliquots of library cultures were washed twice with PBS by centrifuging at 2500xg for 5 minutes and finally resuspended in 8mL PBS.

9.3 Flow cytometry and sorting conditions

Library cultures in PBS were sorted using a Sony SH800 Cell Sorter (Sony, Tokyo, Japan). All flow cytometry and sorting were performed using a 488nm laser and using measurements from a 525/30 band-pass filter. For both methodologies, events were first subjected to size-selection by gating for those that fell into the mode of the FSC-A distribution (~30% of the population).

The threshold sort was conducted using a gate that sorted the size-selected cells that fell into the top 0.5% of the fluorescence distribution using

the excitation and emission listed above. These sorts were used to collect 2000 events, which were immediately plated on solid agar media.

In the sort-seq methodology, each sort consisted of an initial enrichment sort, followed by a purification sort of the sorted culture resulting from the enrichment sort. The enrichment sort used the “yield” sort mode. The enrichment sort collected $4x10^6$ events, which were then used for the purification sort. The purification sort used the same size and fluorescence gates constructed in the yield sort, though the gated portion then represented 40-60% of the population. The purification sort used the “purity” sort mode. The purification sort was allowed to continue until the source material from the enrichment sort was depleted. The purification sort ultimately yielded $8x10^5$-$9x10^5$ total events sorted.

9.4 Preparation of sorted samples for sequencing

The sorted population (~400uL in PBS) was transferred to 50mL of SD lacking tyrosine and uracil and grown for 30 hours to approximately OD/mL = 11-14. The equivalent of 5 ODs of the saturated culture were stored in a glycerol stock. The equivalent of 5 ODs of grown culture were used for a genome preparation using zymolyase (Zymo Research, Irvine, CA) to be analyzed by sequencing.

9.5 Liquid culture measurements of betaxanthin fluorescence

Liquid cultures were grown in 96 well deep-well blocks for 24 hours. Cultures were back diluted 1:100 into fresh media and aliquoted for fluorescence measurements after 8 hours. Optical density (OD) and fluorescence were measured using a Tecan Spark plate reader (Tecan, Zurich, Switzerland). OD was measured at 600nm and betaxanthin fluorescence was measured at an excitation/emission of 485/525nm. Intracellular measurements were performed by aliquoting 100μL of cell culture into v-bottom 96-well plates and centrifuging the plates at 4700rpm, then resuspending pellets in 100μL water before measurement. Extracellular measurements were performed by aliquoting the supernatants from the centrifugation into measurement plates. All measurements were performed using black flat-bottom 96-well plates (Corning, Corning, New York).

9.6 Cas9-mediated markerless gene deletions

Deletions were performed as described in Lee et al(M. E. Lee et al., 2015). Plasmids were constructed containing Cas9 expression cassettes and sgRNA expression cassettes (pJS1324, pBC909, and pJS1602). Repair DNA was designed as two oligonucleotides 58bp in length with a 16bp overlap between them. The remaining 50bp of each oligonucleotide were homologous to either the sequence immediately upstream of the gene's start codon or immediately downstream of the gene's stop codon. Oligonucleotides were transformed along with the Cas9 expression plasmid and the plasmid was selected for by growing on agar media lacking uracil. Selection for the Cas9 plasmid causes cells that have not been edited to suffer toxic double stranded breaks. This serves as a selection for successfully edited cells (M. E. Lee et al., 2015). After selection of edited cells, Cas9 plasmids were removed by selection with 5-fluoroorotic acid.

9.7 Building the transposon mutagenesis cassette

Custom disruption libraries were generated in engineered backgrounds using the transposon shuttle mutagenesis method(Ross-Macdonald et al., 1999). A mu transposon was built containing mu repeats flanking a trimR cassette for selection in *E. coli*, a HygR cassette for selection in *S. cerevisiae*, and a 20bp barcode. The 20bp randomized barcodes were generated by PCR of the transposon vector with primers GL76 and GL77 which contain degenerate bases at the barcode site. These primers also contain SapI restriction enzyme cut sites that enable ligation of the two ends of the PCR product to form a circular plasmid. Linear DNA fragments containing the barcoded transposons were excised from the vector using BamHI and EcoRI restriction enzyme cut sites flanking the mu repeats and gel purified.

In vitro transposition was performed by incubating the excised transposon DNA with mu transposase and an existing plasmid library containing genomic fragments (Jones, Stalker, Humphray, Dunham, & Prelich, 2008a). The resulting transposition reaction was transformed by electroporation into *E. coli* TG1 and grown in LB with kanamycin (50mg/L) and trimethoprim (15mg/L) to select for plasmids from the genome fragment library that contain successful transposition events (the transposons carry trimR). Plasmid DNA amplified in these cultures was extracted using Zymo Mini Prep Kits (Zymo Research, Irvine, CA).

To enrich for plasmids in which the genomic fragments contain the transposition events (as opposed to transpositions in the vector backbone), we used a gateway cloning reaction (Invitrogen) following manufacturer's

protocols to transfer the genomic fragments to a new vector backbone with a different selectable marker (specR). The results of the gateway cloning reaction were transformed by electroporation into *E. coli* TG1 and the plasmid DNA was amplified by growing in LB with trimethoprim (to select for the transposons; 15mg/L) and spectinomycin (to select for the new backbone without transposition events; 50mg/L). Plasmid DNA from this culture was extracted using Zymo Mini Prep Kits to obtain the final *in vitro* transposed genome fragment library. In the new backbone, the genomic fragments were flanked by SceI homing endonuclease cut sites, enabling us to excise a linear fragment containing the genomic fragments to use for efficient integrations into the yeast genome by homologous recombination.

9.8 Transposon and barcode sequencing (TnSEQ and BarSEQ)

Illumina-compatible sequencing libraries were generated to link the random DNA barcodes to the transposon insertion sites in the *in vitro* transposed genome fragment library using the method "TnSeq sequencing library preparation" from Wetmore *et al*.(Wetmore et al., 2015), starting with 1ug of the *in vitro* transposed genome library. The resulting library was sequenced on a HiSeq 2500 system (Illumina). Barcodes were mapped to the transposon insertion site using a Perl script (MapTnSeq.pl) as described in Wetmore *et al*. (Wetmore et al., 2015). Barcodes in enriched strains were identified by PCR amplification of the barcode region using primers FN53 and FN54 followed by Sanger sequencing.

9.9 Production of transposon-based disruption variant libraries

The *in vitro* transposon library was cut using SceI to create a linear fragment containing transposon insertion sites flanked by yeast genomic DNA regions. Transformation of these fragments was performed following the lithium acetate protocol (Gietz & Schiestl, 2007). Transformants were selected on YPD agar plates supplemented with 300μg/mL hygromycin. Plates were scraped and resulting cultures were mixed 1:1 with 50% glycerol. 1mL aliquots of this mixture were stored at -80°C.

9.10 Dopamine measurements

Colonies were picked into 2mL of synthetic complete medium (minus uracil, tyrosine) with 2% glucose. After overnight growth, saturated cultures were back-diluted 100x into 24-well, deep-well blocks with 2mL fresh media.

The cultures were grown in a Multitron shaker (Infors HT, Bottmingen, Switzerland) for 14h at 30°C. Cultures were pelleted, and spent media was removed for further analysis. Dopamine standards were made in spent media of wild type cultures. Samples were mixed 1:1 with *n*-butanol by pipetting up and down 4 times. The organic layer was collected and evaporated, then resuspended in water to prepare for LC/MS.

Samples were loaded onto a Zorbax Eclipse Plus C18 4.6 × 100mm-3.5μm reversed-phase column (Agilent Technologies, Santa Clara, CA) at ~20 °C using a 0.5 ml/min flow rate. Samples were eluted with a constant mixture of 60% water/40% acetonitrile plus 0.1% formic acid over the course of 20 min. Column was washed with linear gradients up to 100% water and 100% acetonitrile over 5 min. MS was carried out using a 6520 Accurate-Mass Q-TOF LC/MS (Agilent Technologies) for fragmentation and mass detection. The system was run with a fragmentor voltage of 100-V and a collision energy of 23 V. Dopamine ion counts were quantified using m/z of 154.086 [M+H]+ and retention time of approximately 8.5 min. Dopamine concentrations were quantified against a calibration curve ranging from 0 to 24mg/L.

9.11 Indigo measurements

Colonies were picked into 0.5mL of synthetic complete medium (minus uracil) with 2% glucose. Cultures were grown overnight at 30°C in a Multitron shaker (Infors HT, Bottmingen, Switzerland). Cultures were normalized to OD 0.4/mL using the same media, then 0.1mL were aliquoted and pelleted. Pellets were resuspended in PBS with 1mM indole and incubated overnight at 30°C. Cultures with indigo precipitate were pelleted and resuspended in 100% DMSO to dissolve the indigo. Indigo was quantified at 560nm absorbance.

9.12 Heme measurements

Total heme was quantified by measuring the fluorescence of protoporphyrin IX (following heme release of iron) (Michener et al., 2012). For each sample, $2.4x10^8$ cells at OD_{600}=0.8 were collected, washed with phosphate buffered saline, and resuspended in 500uL of 20mM oxalic acid in amber tubes. Samples were left at 4°C for 16 hours in a light-tight box. 500uL 2M oxalic acid were added to each sample and the sample was split between two tubes, one of which was heated to 98°C for 30 min. Samples were then centrifuged at 16,000xg for 2 minutes and supernatants of heated and

unheated samples were assayed in a Tecan Spark plate reader (Tecan, Zurich, Switzerland). Measurements were taken at an excitation of 400nm and emission of 620nm. Measurements were compared against dilutions of a hemin standard that was treated to the same heating method.

References

Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., et al. (2010). Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli. Science (New York, N.Y.)*, *330*(6000), 70–74. http://doi.org/10.1126/science.1191652

Alper, H., & Stephanopoulos, G. (2007). Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metabolic Engineering*, *9*(3), 258–267. http://doi.org/10.1016/j.ymben.2006.12.002

Arnow, L. E. (1937). Colorimetric determination of the components of 3, 4-dihydroxyphenylalanine-tyrosine mixtures. *Journal of Biological Chemistry*, *118*, 5321–5537.

Bao, Z., HamediRad, M., Xue, P., Xiao, H., Tasan, I., Chao, R., et al. (2018). Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. *Nature Biotechnology*, *36*(6), 505–508. http://doi.org/10.1038/nbt.4132

Biggs, B. W., Lim, C. G., Sagliani, K., Shankar, S., Stephanopoulos, G., De Mey, M., & Ajikumar, P. K. (2016). Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in *Escherichia coli. Proceedings of the National Academy of Sciences*, *113*(12), 201515826–3214. http://doi.org/10.1073/pnas.1515826113

Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., & Boeke, J. D. (1998). Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)*, *14*(2), 115–132. http://doi.org/10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2

Cambray, G., Guimaraes, J. C., & Arkin, A. P. (2018). Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*. *Nature Biotechnology*, *36*(10), 1005. http://doi.org/10.1038/nbt.4238

Chang, M. C. Y., Eachus, R. A., Trieu, W., Ro, D.-K., & Keasling, J. D. (2007). Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. *Nature Chemical Biology*, *3*(5), 274–277. http://doi.org/10.1038/nchembio875

Chen, K., Huang, X., Kan, S. B. J., Zhang, R. K., & Arnold, F. H. (2018). Enzymatic construction of highly strained carbocycles. *Science (New York, N.Y.)*, *360*(6384), 71–75. http://doi.org/10.1126/science.aar4239

Coradetti, S. T., Pinel, D., Geiselman, G. M., Ito, M., Mondo, S. J., Reilly, M. C., et al. (2018). Functional genomics of lipid metabolism in the oleaginous yeast *Rhodosporidium toruloides*. *eLife*, *7*, 283. http://doi.org/10.7554/eLife.32110

DeLoache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J. J., & Dueber, J. E. (2015). An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nature Chemical Biology*, *11*(7), 465–471. http://doi.org/10.1038/nchembio.1816

Dietrich, J. A., Yoshikuni, Y., Fisher, K. J., Woolard, F. X., Ockey, D., McPhee, D. J., et al. (2009). A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous P450(BM3). *ACS Chemical Biology*, *4*(4), 261–267. http://doi.org/10.1021/cb900006h

Eason, R. G., Pourmand, N., Tongprasit, W., Herman, Z. S., Anthony, K., Jejelowo, O., et al. (2004). Characterization of synthetic DNA bar codes in *Saccharomyces cerevisiae* gene-deletion strains. *Proceedings of the National Academy of Sciences*, *101*(30), 11046–11051. http://doi.org/10.1073/pnas.0403672101

Edgar, S., Li, F.-S., Qiao, K., Weng, J.-K., & Stephanopoulos, G. (2016). Engineering of Taxadiene Synthase for Improved Selectivity and Yield of a Key Taxol Biosynthetic Intermediate. *ACS Synthetic Biology*, *6*(2), 201–205. http://doi.org/10.1021/acssynbio.6b00206

Foster, P. L. (1991). In vivo mutagenesis. *Methods in Enzymology*, *204*, 114–125.

Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., & Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science (New York, N.Y.)*, *349*(6252), 1095–1100. http://doi.org/10.1126/science.aac9373

Gelperin, D. M. (2005). Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes & Development*, *19*(23), 2816–2826. http://doi.org/10.1101/gad.1362105

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., et al. (2003). Global analysis of protein expression in yeast. *Nature*, *425*(6959), 737–741. http://doi.org/10.1038/nature02046

Giaever, G., & Nislow, C. (2014). The yeast deletion collection: a decade of functional genomics. *Genetics*, *197*(2), 451–465. http://doi.org/10.1534/genetics.114.161620

Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, *2*(1), 31–34. http://doi.org/10.1038/nprot.2007.13

Gold, N. D., Gowen, C. M., Lussier, F.-X., Cautha, S. C., Mahadevan, R., & Martin, V. J. (2015). Metabolic engineering of a tyrosine-overproducing yeast platform using targeted metabolomics. *Microbial Cell Factories*, *14*(1), 73. http://doi.org/10.1186/s12934-015-0252-2

Goodman, D. B., Church, G. M., & Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science (New York, N.Y.)*, *342*(6157), 475–479. http://doi.org/10.1126/science.1241934

Gorsich, S. W., Dien, B. S., Nichols, N. N., Slininger, P. J., Liu, Z. L., & Skory, C. D. (2006). Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes ZWF1, GND1, RPE1, and TKL1 in *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology*, *71*(3), 339–349. http://doi.org/10.1007/s00253-005-0142-3

Grewal, P. S., Modavi, C., Russ, Z. N., Harris, N. C., & Dueber, J. E. (2018). Bioproduction of a betalain color palette in Saccharomyces cerevisiae. *Metabolic Engineering*, *45*, 180–188. http://doi.org/10.1016/j.ymben.2017.12.008

Hagel, J. M., & Facchini, P. J. (2013). Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant and Cell Physiology*, *54*(5), 647–672. http://doi.org/10.1093/pcp/pct020

Hartmann, M., Schneider, T. R., Pfeil, A., Heinrich, G., Lipscomb, W. N., & Braus, G. H. (2003). Evolution of feedback-inhibited beta /alpha barrel isoenzymes by gene duplication and a single mutation. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(3), 862–867. http://doi.org/10.1073/pnas.0337566100

Hawkins, K. M., & Smolke, C. D. (2008). Production of benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*. *Nature Chemical Biology*, *4*(9), 564–573. http://doi.org/10.1038/nchembio.105

Ho, C. H., Magtanong, L., Barker, S. L., Gresham, D., Nishimura, S., Natarajan, P., et al. (2009). A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nature Biotechnology*, *27*(4), 369–377. http://doi.org/10.1038/nbt.1534

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

Jensen, K., & Møller, B. L. (2010). Plant NADPH-cytochrome P450 oxidoreductases. *Phytochemistry*, *71*(2-3), 132–141. http://doi.org/10.1016/j.phytochem.2009.10.017

Jones, G. M., Stalker, J., Humphray, S., Dunham, I., & Prelich, G. (2008a, March). A systematic library for comprehensive overexpression screens in *Saccharomyces cerevisiae*. Retrieved August 27, 2015, from http://www.nature.com/nmeth/journal/v5/n3/pdf/nmeth.1181.pdf

Jones, G. M., Stalker, J., Humphray, S., West, A., Cox, T., Rogers, J., et al. (2008b). A systematic library for comprehensive overexpression screens in, *5*(3), 239–241. http://doi.org/10.1038/NMETH.1181

Jung, S. T., Lauchli, R., & Arnold, F. H. (2011). Cytochrome P450: taming a wild type enzyme. *Current Opinion in Biotechnology*, *22*(6), 809–817. http://doi.org/10.1016/j.copbio.2011.02.008

Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., et al. (2013). Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(34), 14024–14029. http://doi.org/10.1073/pnas.1301301110

Krainer, F. W., Capone, S., Jäger, M., Vogl, T., Gerstmann, M., Glieder, A., et al. (2015). Optimizing cofactor availability for the production of recombinant heme peroxidase in *Pichia pastoris*. *Microbial Cell Factories*, *14*(1), 4. http://doi.org/10.1186/s12934-014-0187-z

Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, *4*(9), 975–986. http://doi.org/10.1021/sb500366v

Leonard, E., & Koffas, M. A. G. (2007). Engineering of Artificial Plant Cytochrome P450 Enzymes for Synthesis of Isoflavones by *Escherichia coli*. *Applied and*

Environmental Microbiology, *73*(22), 7246–7251. http://doi.org/10.1128/AEM.01411-07
LeProust, E. M., Peck, B. J., Spirin, K., McCuen, H. B., Moore, B., Namsaraev, E., & Caruthers, M. H. (2010). Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Research*, *38*(8), 2522–2540. http://doi.org/10.1093/nar/gkq163
Lewis, J. C., Mantovani, S. M., Fu, Y., Snow, C. D., Komor, R. S., Wong, C.-H., & Arnold, F. H. (2010). Combinatorial alanine substitution enables rapid optimization of cytochrome P450BM3 for selective hydroxylation of large substrates. *ChemBioChem*, *11*(18), 2502–2505. http://doi.org/10.1002/cbic.201000565
Lewis, N. E., Cho, B.-K., Knight, E. M., & Palsson, B. Ø. (2009). Gene expression profiling and the use of genome-scale in silico models of *Escherichia coli* for analysis: providing context for content., *191*(11), 3437–3444. http://doi.org/10.1128/JB.00034-09
Mansell, T. J., Warner, J. R., & Gill, R. T. (2013). Trackable multiplex recombineering for gene-trait mapping in *E. coli. Methods in Molecular Biology (Clifton, N.J.)*, *985*(Chapter 12), 223–246. http://doi.org/10.1007/978-1-62703-299-5_12
McIntosh, J. A., Farwell, C. C., & Arnold, F. H. (2014). Expanding P450 catalytic reaction space through evolution and engineering. *Current Opinion in Chemical Biology*, *19*, 126–134. http://doi.org/10.1016/j.cbpa.2014.02.001
Michener, J. K., Nielsen, J., & Smolke, C. D. (2012). Identification and treatment of heme depletion attributed to overexpression of a lineage of evolved P450 monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(47), 19504–19509. http://doi.org/10.1073/pnas.1212287109
Minami, H., Kim, J.-S., Ikezawa, N., Takemura, T., Katayama, T., Kumagai, H., & Sato, F. (2008). Microbial production of plant benzylisoquinoline alkaloids. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(21), 7393–7398. http://doi.org/10.1073/pnas.0802981105
Oh, J., Fung, E., Price, M. N., Dehal, P. S., Davis, R. W., Giaever, G., et al. (2010). A universal TagModule collection for parallel genetic analysis of microorganisms. *Nucleic Acids Research*, *38*(14), e146. http://doi.org/10.1093/nar/gkq419
Özaydın, B., Burd, H., Lee, T. S., & Keasling, J. D. (2013). Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production. - PubMed - NCBI. *Metabolic Engineering*, *15*, 174–183. http://doi.org/10.1016/j.ymben.2012.07.010
Rasool, S., & Mohamed, R. (2015). Plant cytochrome P450s: nomenclature and involvement in natural product biosynthesis. *Protoplasma*, 1–13. http://doi.org/10.1007/s00709-015-0884-4
Renault, H., Bassard, J.-E., Hamberger, B., & Werck-Reichhart, D. (2014). Cytochrome P450-mediated metabolic engineering: current progress and future challenges. *Current Opinion in Plant Biology*, *19*, 27–34. http://doi.org/10.1016/j.pbi.2014.03.004

Ross-Macdonald, P., Sheehan, A., Friddle, C., Roeder, G. S., & Snyder, M. (1999). Transposon mutagenesis for the analysis of protein production, function, and localization. *Methods in Enzymology*, *303*, 512–532.
Smith, A. M., Heisler, L. E., Mellor, J., Kaper, F., Thompson, M. J., Chee, M., et al. (2009). Quantitative phenotyping via deep barcode sequencing. *Genome Research*, *19*(10), 1836–1842. http://doi.org/10.1101/gr.093955.109
Stierle, A., Strobel, G., & Stierle, D. (1993). Taxol and taxane production by *Taxomyces andreanae*, an endophytic fungus of Pacific yew. *Science (New York, N.Y.)*, *260*(5105), 214–216.
Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, *460*(7257), 894–898. http://doi.org/10.1038/nature08187
Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A., & Gill, R. T. (2010). Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nature Biotechnology*, *28*(8), 856–862. http://doi.org/10.1038/nbt.1653
Wetmore, K. M., Price, M. N., Waters, R. J., Lamson, J. S., He, J., Hoover, C. A., et al. (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio*, *6*(3), e00306–15. http://doi.org/10.1128/mBio.00306-15
Whitehouse, C. J. C., Bell, S. G., Yang, W., Yorke, J. A., Blanford, C. F., Strong, A. J. F., et al. (2009). A highly active single-mutation variant of P450BM3 (CYP102A1). *ChemBioChem*, *10*(10), 1654–1656. http://doi.org/10.1002/cbic.200900279
Withers, S. T., & Keasling, J. D. (2007). Biosynthesis and engineering of isoprenoid small molecules. *Applied Microbiology and Biotechnology*, *73*(5), 980–990. http://doi.org/10.1007/s00253-006-0593-1
Yoshikawa, K., Tanaka, T., Ida, Y., Furusawa, C., Hirasawa, T., & Shimizu, H. (2011). Comprehensive phenotypic analysis of single-gene deletion and overexpression strains of *Saccharomyces cerevisiae*. *Yeast (Chichester, England)*, *28*(5), 349–361. http://doi.org/10.1002/yea.1843
Zhang, M. M., Wang, Y., Ang, E. L., & Zhao, H. (2016). Engineering microbial hosts for production of bacterial natural products. *Natural Product Reports*, *33*(8), 963–987. http://doi.org/10.1039/c6np00017g
Zhao, J., Bao, X., Wang, S., Lu, S., Sun, J., & Yang, X. (2017). In Situ Fluorogenic and Chromogenic Reactions for the Sensitive Dual-Readout Assay of Tyrosinase Activity. *Analytical Chemistry*, *89*(19), 10529–10536. http://doi.org/10.1021/acs.analchem.7b02739

Appendices

A. Plasmid List

All plasmids were constructed using the yeast toolkit mo-clo system previously described by our lab (M. E. Lee et al., 2015). All plasmid backbones contained ColE1 replication origins.

Name	Plasmid contents
pJS1051	For integration into S. cerevisiae following linearization by NotI: pTDH3-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DOD-tADH1-pPGK1-ARO4_K229L-tPGK1; KanR *E. coli* marker; yeast URA3 locus homology arms, flanked by NotI restriction enzyme sites
pJS1475	For integration into S. cerevisiae following linearization by NotI: pRPL18B-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DOD-tADH1-pPGK1-ARO4_K229L-tPGK1; KanR *E. coli* marker; yeast URA3 locus homology arms, flanked by NotI restriction enzyme sites
pPSG533	Spacer sequence; KanR *E. coli* marker; *S. cerevisiae* HIS3 expression cassette; *S. cerevisiae* Cen6 replication origin
pZNR1441	pCCW12-HEM2-tENO2-pTDH3-HEM12-tSSA1-pTEF1-HEM3-tTDH1; KanR *E. coli* marker; *S. cerevisiae* HIS3 expression cassette; *S. cerevisiae* Cen6 replication origin
pJS1324	pPGK1-Cas9-tPGK1-sgRNA targeting QDR2; KanR *E. coli* marker; yeast URA3 expression cassette; *S. cerevisiae* Cen6 replication origin
pBC909	pPGK1-Cas9-tPGK1-sgRNA targeting YOR1; KanR *E. coli* marker; yeast URA3 expression cassette; *S. cerevisiae* Cen6 replication origin
pJS1602	pPGK1-Cas9-tPGK1-sgRNA targeting HMX1; KanR *E. coli* marker; yeast URA3 expression cassette; *S. cerevisiae* Cen6 replication origin

B. Strain List

All strains were constructed using the base strain BY4741: *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0.* Purchased deletion collection strains were derived from the same background. *In vitro* transposon insertion library was constructed using genomic DNA isolated from this strain.

Name	Genotype
yJS1051	*ura3::pTDH3-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DOD-tADH1-pPGK1-ARO4_K229L-tPGK1*

yJS1067	*pdr8Δ0; ura3::pTDH3-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DOD-tADH1-pPGK1-ARO4_K229L-tPGK1*
yJS1068	*qdr2Δ0; ura3::pTDH3-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DOD-tADH1-pPGK1-ARO4_K229L-tPGK1*
yJS1159	*qdr2Δ0; ura3::pRPL18B-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DOD-tADH1-pPGK1-ARO4_K229L-tPGK1*
yJS1256	*yor1Δ0; qdr2Δ0; ura3::pRPL18B-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DOD-tADH1-pPGK1-ARO4_K229L-tPGK1*
yJS1257	*yor1Δ0; ura3::pRPL18B-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DOD-tADH1-pPGK1-ARO4_K229L-tPGK1*
yJS1337	*hmx1Δ0; ura3::pRPL18B-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DOD-tADH1-pPGK1-ARO4_K229L-tPGK1*
yJS1221	*ura3::pTDH3-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DODC-tADH1-pPGK1-ARO4_K229L-tPGK1*
yJS1343	*hmx1Δ0; ura3::pTDH3-CYP76AD1_H3H_W13L_F309L-tTDH1-pCCW12-DODC-tADH1-pPGK1-ARO4_K229L-tPGK1*
yJS1372	*ura3::pTDH3-CYP102A1_G4-mRuby2-tADH1* plasmid pPSG533
yJS1373	*hmx1Δ0; ura3::pTDH3-CYP102A1_G4-mRuby2-tADH1* plasmid pPSG533
yJS1558	plasmid pPSG533
yJS1559	*hmx1Δ0;* plasmid pPSG533
yJS1560	plasmid pZNR1441

C. Oligo List

Name	Purpose	Sequence
GL76	Generate barcode	CTTAGCTCTTCACGTACGCTGCAGGTCGAC
GL77	Generate barcode	GAATGCTCTTCAACGNNNNNNNNNNNNNNNNNNNAGAGACCTCGTGGACATC
FN53	Sequence barcode	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGTCGACCTGCAGCGTACG
FN54	Sequence barcode	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT
HY56	QDR2 repair DNA F	TGCTGGTCATTTTAGTAGAAACTCTGCTCTCAAACTTGAGTACTGCAACGGAAGTAAA
HY57	QDR2 repair DNA R	GTGGAGCGATCAAAAGGAACATTTTCCTTTGATTCAAGAAGCTTTACTTCCGTTGCAG
HU80	YOR1 repair DNA F	CGCTAGAATTGAATTTGCCTTATCTTTTCAGCCGTTTTTGGCTGGGTTCCCGCGGCTT
HU81	YOR1 repair DNA R	GACCAGGCAATTGTAATACATAAGTCAACAAAACACCAACTGAAGCCGCGGGAACCCA

IE66	HMX1 repair DNA F	GcacaatataacacagcatatatacacacacacacataAAATA ACCGCAAAACGTATaaa
IE67	HMX1 repair DNA R	ATATTATTTCATGTATATATTATGTTTGTATTTAG ACttttttttttATACGTTTTGCGG