High-throughput methods of studying human cytochrome P450 activity in *Saccharomyces cerevisiae*

Clara J. Amorosi

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Maitreya J. Dunham, Chair

Douglas M. Fowler

Allan E. Rettie

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

High-throughput methods of studying human cytochrome P450 activity in *Saccharomyces cerevisiae*

Clara J. Amorosi

Chair of the Supervisory Committee:
Maitreya J. Dunham, Professor
Department of Genome Sciences

Genetic variation in cytochrome P450 enzymes leads to inter-individual variability in drug metabolism, while lack of functional annotation prevents most variants from being used to inform drug choice and dosing. To measure P450 activity in a high-throughput manner, we developed a yeast-based activity assay. *S. cerevisiae* has a long history of use as a heterologous system of mammalian CYP expression, but we improved on this with targeted strain engineering and screening of natural isolates. In the process, we identified a sake strain with much higher human CYP expression levels than the laboratory strain. To measure P450 activity in yeast cells, we developed an assay using activity-based protein profiling with novel P450 activity-based probes to label a pooled population of variants in an activity-dependent manner. We extended this approach to a particularly important pharmacogene: *CYP2C9*, genetic variation in which affects the efficacy of warfarin, phenytoin, and other drugs. We coupled our yeast activity assay

with fluorescence-activated cell sorting and high throughput sequencing, and generated activity scores for 6,142 single missense variants of CYP2C9. Strikingly, 65% of missense variants have significantly decreased activity suggesting altered drug metabolism *in vivo*. With collaborators, we also performed a second deep mutational scan of CYP2C9 in a human cell line, measuring variant abundance for 6,370 single missense variants, revealing that stability plays a large role in CYP2C9 function. Our yeast activity assay can be extended to other CYP enzymes, and will lead to advances in adverse drug response prevention by providing clinical guidance for patients carrying both currently known and yet-to-be discovered alleles of *CYP2C9*.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I have been blessed with wonderful science teachers and mentors throughout my academic career. Thank you to Dewey Moody, Ellen Reimer, Eric Muhs, and Penny Pagels for making high school science fun, not an easy task. At Harvey Mudd, I was mentored by Dan Stoebel, who taught me molecular biology basics and the joy of benchwork, and set the bar high for future mentors. At UW, I was incredibly fortunate to join the lab of Maitreya Dunham, who has been a joy to work with. Her passion for science has gotten me through many failed experiments. Maitreya has gone above and beyond to help me pursue my research interests and has created an amazingly supportive lab environment. Thank you to my committee members, Doug Fowler, Allan Rettie, Jay Shendure, and Phil Green, who have provided crucial advice and support throughout my time in graduate school. Thanks to the many official and unofficial mentors I've had throughout graduate school, especially Elyse Hope, Elisa Wong, Kate Sitko, and Melissa Chiasson, who were always willing to take the time to teach me something new. Thanks to the amazing scientists in my cohort, especially Katherine Xue, Serena Liu, Emily Killingbeck, and Claudia Espinoza, for providing countless hours of support and commiseration. And finally, thanks to my parents, Gayle and Steve Amorosi, for their unending support from day 1, and to my partner, Adam Gilbert, for being my rock throughout grad school.

Chapter 1. Introduction

1.1 History and clinical implementation of pharmacogenomics

The goal of pharmacogenomics (PGx) is to understand the genetic sources of inter-individual variation in drug response. While the field of pharmacogenetics has existed for over 50 years (Motulsky 1957), the clinical implementations of the knowledge gained has lagged behind the discovery of genetic interactions with many drugs. In the US, adverse drug reactions (ADRs) are a leading cause of hospitalization and death (Lazarou et al. 1998). ADRs can be caused by overdoses, harmful drug interactions, and genetic factors. 30% of ADRs are predicted to be caused by inter-individual variability in drug metabolizing enzymes and other drug related genes (Budnitz et al. 2006). As of 2015, germline genetic variants had been identified in 14 genes that affect the metabolism of about 80 medications that are actionable in the clinic (Relling and Evans 2015). Similarly, 18% of all prescriptions in the U.S. were for drugs with actionable pharmacogenetics (Relling and Evans 2015).

Even when there is compelling evidence for a gene-drug pair that meets the threshold for clinical action, PGx screening and genotype-guided dosing is not incorporated into most routine clinical care. The exception to this is the HIV antiviral abacavir, where roughly 5% of patients will develop potentially life-threatening hypersensitivity reactions due to a polymorphism in *HLA-B*, so genetic screening for *HLA-B*57:01* is now considered routine clinical practice before prescribing abacavir (Martin and Kroetz 2013). In other cases when PGx is incorporated into clinical care, it is often in the form of single-gene tests that are ordered reactively when prescribing a medication. Such a gene-by-gene approach is slow and costly, and in the future

PGx screening will likely be performed using multi-gene panels that include many pharmacogenes that cover most PGx high-risk drugs (Dunnenberger et al. 2015). Preemptive (pre-prescription) genotyping approaches have been implemented at a handful of institutions and have shown promise (Bielinski et al. 2014, Gottesman et al. 2013, Pulley et al. 2012, Luzum et al. 2017). Furthermore, a pilot study that prospectively genotyped five pharmacogenes found that a remarkable 99% of the patients carried an actionable PGx variant in at least one gene (Ji et al. 2016), indicating the usefulness of a preemptive PGx gene panel.

Incorporating PGx testing in routine patient care has the potential to be cost-effective, improve treatment, and decrease ADRs. ADRs are estimated to cost U.S. hospitals up to 4 billion dollars annually (Lazarou et al. 1998), though a more recent review estimated that the total impact and management of ADRs costs the U.S. up to 30 billion annually (Sultana et al. 2013), and an estimated 4.3% of hospital admissions are due to preventable ADRs (Winterstein et al. 2002). In general, randomized control trials of genotype-guided dosing have shown that implementing genotype-guided dosing was effective at improving patient outcomes, though this depended on the trial and gene-drug pair (Goulding et al. 2015). There is still a large gap between the promise of personalized medicine and its implementation in routine clinical care.

To help bridge this gap, the Clinical Pharmacogenetics Implementation Consortium (CPIC) reviews *in vitro* and *in vivo* evidence and publishes clinical functional recommendations for actionable gene-drug pairs (Relling and Klein 2011). These recommendations are intended to help clinicians use available genetic test results to optimize drug therapy by adjusting dosage or changing medication, with the assumption that preemptive genotyping will become more widespread. Gene-drug pairs are clinically actionable if they can lead to a severe clinical

consequence (i.e. potential lethality for abacavir) and have documented alleles that impact drug metabolism. For example, in the treatment of leukemia, genetic polymorphisms in *TPMT* and *NUDT15* can lead to thiopurine-induced hematopoietic toxicity, and genotype-guided thiopurine dosing guidelines have been put forward by CPIC (Relling et al. 2019). Ultimately, however, putting forward genotype-guided dosing recommendations for these and other gene-drug pairs is just one of the steps to implementing precision medicine in routine clinical care.

1.2 Biochemistry of cytochrome P450 enzymes

Cytochrome P450s, or CYPs, are some of the most important pharmacogenes due to the large number of drugs metabolized by these enzymes (Zanger and Schwab 2013). CYPs are a large family of heme proteins responsible for metabolizing a wide range of drugs and xenobiotics, and play a central role in Phase I drug metabolism, in which substrates undergo oxidation, hydrolysis, or reduction in preparation for further processing (Ionescu and Caira 2005). CYPs require a cytochrome P450 reductase (CPR), which transfers electrons from NADPH to the CYP. Cytochrome b5 also can play a role in CYP reduction. Humans have 57 functional CYP genes (Nelson et al. 2004), which are classified according to sequence identity into 18 families and 44 subfamilies. Families CYP1, CYP2 and CYP3 are involved in biotransformation of most exogenous substances (Zanger and Schwab 2013), and 12 CYP enzymes in these families are collectively responsible for 75% of all known drug oxidation reactions (Evans and Relling 1999).

Cytochrome P450 enzymes are primarily found in liver cells but can also be found throughout the body. Quantification of relative CYP expression levels based on mass spectrometry showed that CYP2E1 is present at 25-33% of total hepatic P450 protein, CYP2C9

at ~25%, CYP3A4 at ~15%, CYP1A2 at ~10%, and CYP3A5 at 1-10% (Gröer at al. 2014, Zhang et al. 2016). However, many liver P450s are inducible, and relative expression levels can also vary with age, sex, polymorphism, and many other environmental and regulatory factors (Zanger and Schwab 2013).

1.3 Genetic diversity in cytochrome P450s: Rare and common variants

CYP genes are highly polymorphic and have a large number of single nucleotide variants (SNVs) and copy number variants. CYP alleles are named according to a star (*) nomenclature where each haplotype is assigned a star number (*1 is assigned to the wildtype or consensus allele) (Sim and Ingelman-Sundberg 2010). In *CYP2C9*, *CYP2C19*, and *CYP2D6,* the high number of gene variants found in different populations results in significant inter-individual variability in drug metabolism (Ji et al. 2016, Zhou et al. 2017), and these are considered some of the most clinically actionable CYPs.

Variants in *CYP2C9* are highly population-specific, with the *2 missense allele most common in Europeans (MAF = 11.7%), the *3 missense allele most common in South Asians (MAF = 11.3%), and the *8 and *9 missense alleles almost exclusively found in Africans (MAF of 5.6% and 7.5%, respectively) (Zhou et al. 2017). For *CYP2C19*, the *2 splicing variant is most common in East and South Asians (MAF of 31.0% and 34.0% respectively), while the *17 promoter variant is most common in Europeans, Africans, and admixed Americans (MAF of 22.4%, 23.5%, and 12.0%, respectively) (Zhou et al. 2017). Other *CYP2C19* alleles are also highly population-specific. In contrast to *CYP2C9* and *CYP2C19* where alleles are dominated by SNVs, the *CYP2D6* gene locus has many alleles with large structural rearrangements due to its proximity to pseudogene *CYP2D7* (Yang et al. 2017). For *CYP2D6*, the most prevalent

haplotype is the *2 missense allele (MAF of 34.3%, 26.7%, and 36.2% in Europeans, Africans, and South Asians, respectively), followed by the *4 splice variant allele (MAF of 15.5%, 11.9%, and 11.6% in Europeans, Africans, and South Asians, respectively) (Zhou et al. 2017). Additionally, *CYP2D6* gene duplications are common, showing frequencies of up to 29% in certain African populations (Aklillu et al. 1996).

In addition to the common alleles, recent population sequencing efforts have given a better picture of rare variation present in P450 genes. A study that analyzed sequencing from the NHLBI and 1000 Genomes project found that the vast majority (82%) of SNVs in coding regions of P450 genes was rare (MAF < 0.1%) (Fujikura et al. 2015). Further sequencing of 6,000 individuals with a targeted PGx gene panel found that 10% of individuals had a previously unknown variant in at least one CYP (Gordon et al. 2014), emphasizing that rare alleles can be "collectively common" *en masse*.

The PharmVar Consortium (Gaedigk et al. 2018) shares a repository of PGx haplotype and allele information, listing 62, 38, and 139 different star alleles for *CYP2C9*, *CYP2C19*, and *CYP2D6*, respectively. Clinical functional status designations (increased function, normal function, decreased function, and no function) curated by the Clinical Pharmacogenomics Implementation Consortium (CPIC) are available for 36, 26, and 58 of the star alleles for CYP2C9, CYP2C19, and CYP2D6, respectively (Relling and Klein 2011). CPIC has published guidelines for how to implement pharmacogenomic information in a clinical setting for over 50 drugs, including warfarin.

Warfarin is a widely prescribed oral anticoagulant with a narrow therapeutic window. Patients taking warfarin must be monitored to avoid hemorrhage or blood clots, and dose is often

altered before reaching a stable dose. Warfarin is metabolized by CYP2C9 and targets vitamin K oxidoreductase in the vitamin K recycling pathway, encoded by *VKORC1*. *CYP2C9* genotype determines 15-20% of the variation in warfarin dose (Duconge et al. 2009), while *VKORC1* polymorphisms contribute to an estimated 25% of variation in warfarin dose (Owen et al. 2010). *CYP2C9* variants can lead to warfarin sensitivity (daily warfarin dose < 1.5 mg), while variation in *VKORC1* noncoding and coding sequence can cause warfarin resistance (daily warfarin dose > 6 mg) or sensitivity (Rieder et al. 2005, Yuan et al. 2005). Genotype-guided warfarin dosing can improve patient treatment in some situations (Pirmohamed 2013), but relies on alleles of known function to guide dosing decisions. Therefore, there is a need for additional functional annotation of these and other P450 variants, but efforts to characterize known variants are hampered by the low throughput of CYP biochemical assays, which often involve LC-MS.

1.4 Existing methods for testing P450 activity *in vitro* and *in vivo*

A variety of methods have been developed in order to determine CYP substrate metabolic profiles, potential drug-drug interactions, and the impact of variants on drug metabolism. Two main approaches are i) LC-MS for metabolite detection, and ii) fluorimetric methods that measure the formation of highly fluorescent metabolites from fluorogenic CYP substrates. In general, the first approach is limited in sample throughput and is generally performed on cellular fractions, while the second is amenable to multiplexing in 96-well plates and can be performed on cellular fractions or whole cells (Donato and Gómez-Lechón 2013). Probe substrates have been identified for most common human liver CYPs for both LC-MS assays (Walsky and Obach 2004) and fluorimetric assays (Donato et al. 2004, Foti et al. 2010). Because there is often substrate overlap between CYPs (that is, different CYP isoforms may metabolize the same

substrates), it is frequently necessary to use a recombinant system that expresses only one human CYP in order to determine the contribution of that individual CYP to substrate turnover.

Despite this, the gold standard P450 biochemical assay is metabolite quantification using LC-MS in human liver microsomes (the cell fraction containing the endoplasmic reticulum). CYP2C9 has well-documented LC-MS assays using (S)-warfarin and (S)-flurbiprofen (Mosher et al. 2008), while (S)-mephenytoin can be used for CYP2C19 (Wada et al. 2008) and desipramine for CYP2D6 (Hanson et al. 2010). One final method of measuring P450 activity is measuring blood plasma concentrations of CYP metabolites, but these methods are extremely low throughput and only amenable to testing CYP variants at high enough population frequency to find homozygotes, the ideal case (Kirchheiner et al. 2003, Stearns et al. 2003).

1.5 History of CYP2C9 pharmacogenetics research

Though there are currently hundreds of known substrates for human CYPs, determining which substrate(s) are metabolized by which CYP isoform(s) was a relatively recent effort and is still ongoing for many CYPs. Often, the discovery of a specific P450 isoform was due to poor metabolism phenotypes observed in individuals, as was the case for CYP2C9. In 1985, a case study of a tolbutamide poor metabolizer showed evidence that tolbutamide was metabolized by a P450 isoform distinct from those that metabolize debrisoquine (metabolized by CYP2D6) and theophylline (metabolized by CYP1A2) (Miners et al. 1985). The enzyme responsible for tolbutamide hydroxylation was later identified as CYP2C9, and was also shown to be responsible for S-warfarin 7-hydroxylation (Rettie et al. 1992) and phenytoin 4-hydroxylation (Veronese et al. 1991).

Alleles of *CYP2C9* were soon identified that were responsible for poor metabolizer phenotypes. An individual homozygous for I359L (*3) was confirmed as a tolbutamide poor metabolizer (Sullivan-Klose et al. 1996), and a separate individual also homozygous for *3 was shown to have diminished S-warfarin clearance leading to warfarin sensitivity (Steward et al. 1997). Following this, targeted cDNA sequencing studies determined the frequency of major *CYP2C9* alleles *2 and *3 in Caucasian populations (Stubbins et al. 1996). In a retrospective cohort study, these alleles were shown to cause adverse warfarin outcomes (overanticoagulation and bleeding events), providing a direct link between patient genotype and increased risk of a serious or life threatening bleeding event (Higashi et al. 2002). Around this time, the crystal structure of CYP2C9 bound to warfarin (Williams et al. 2003) and flurbiprofen (Wester et al. 2004) was published, providing insight into CYP2C9 structure-function. In particular, the flurbiprofen-bound structure highlighted the key role of Arg-108 in determining CYP2C9's preference for acidic substrates (Wester et al. 2004). In addition to well-documented CYP2C9 substrates such as warfarin, tolbutamide, and phenytoin, CYP2C9 metabolizes a large number of other substrates, including a newly-approved multiple sclerosis drug, siponimod. This drug is primarily metabolized by CYP2C9, and is contraindicated for treatment in *3/*3 homozygotes due to risk of bradycardia (Novartis 2019), showing the continued importance of CYP2C9. In these CYP2C9 studies, patient genotyping/phenotyping was crucial for determining the CYP2C9 pharmacogenetics. However, acquiring patient samples is challenging and many studies have turned to heterologous CYP expression systems to characterize variant function.

1.6 Using yeast as a heterologous CYP expression system

When investigating the function of CYP alleles, testing in a physiologically relevant system such as human liver microsomes or human hepatocytes remains the gold standard, but can be challenging with low-frequency alleles. Thus, assays in a variety of recombinant heterologous expression systems play an important role in characterizing new alleles. Human CYPs have been expressed in *Escherichia coli*, yeast, baculoviruses, and mammalian cell lines (Hiratsuka 2011). *E. coli* are used due to low cost and high yield but target genes require N-terminal modification to express mammalian CYPs (Barnes et al. 1991). Yeast, specifically baker's yeast *Saccharomyces cerevisiae*, are used due to low cost, genetic tractability, highly scalable culture volumes, and do not require CYP modification for expression. Baculovirus expression in insect cells can produce very high yields of 300 - 1,000 pmol/g (Asseffa et al. 1989, Gonzalez et al. 1991) and are often used for commercial microsome preparations (Corning Inc.), but are also more expensive. Finally, CYP expression in cell lines (most commonly monkey kidney COS cell lines and human kidney HEK293 cell lines) provides a more relevant expression system at the cost of low expression (Minowa et al. 1990) and higher expense. In a study that characterized expression of a bovine CYP across different expression systems, baculovirus and *E. coli* were expressed at the highest level, followed by yeast microsomes with intermediate expression and COS microsomes with lowest expression (Waterman 1993).

As previously mentioned, yeast are ideally suited for expressing recombinant CYPs because of their genetic tractability, ease of culturing, and ability to accurately localize recombinant CYPs with no additional modifications. Because of this, yeast have been used as a system for expressing recombinant CYPs for over 30 years (Oeda et al. 1985). Early yeast

systems expressed CYP cDNAs of interest from a plasmid (Oeda et al. 1985, Shibata et al. 1990), and had adequate expression for spectral and catalytic assays when starting with large culture volumes. However, subsequent yeast expression systems added in human accessory genes such as CPR and b5, and even constructed CYP-CPR fusions (Urban et al. 1993, Urban et al. 1994, Yabusaki 1995). These modifications increased the activity of recombinant CYPs in yeast. Currently, there is no universally accepted optimal yeast system for heterologous CYP expression, but most reported systems coexpress CYP and CPR to improve CYP activity (Pompon et al. 1994). Yeast microsomes have been used to characterize the function of many human CYP alleles (Sullivan-Klose et al. 1996, Takanashi et al. 2000, Wennerholm et al. 2001). For higher-throughput variant characterization, new yeast strain modifications may be necessary to improve CYP expression.

1.7 Introduction to deep mutational scanning in pharmacogenes

In the past decade, innovations in high throughput sequencing have allowed the development of 'multiplexed assays of variant effects' (MAVEs), where the function of hundreds to thousands of sequence variants can be measured simultaneously (Gasperini et al. 2016). The two main types of MAVEs are deep mutational scanning (DMS) and massively parallel reporter assays. Deep mutational scanning is a high-throughput method for probing protein function by applying a selective pressure to a library of protein variants, enriching for variants with high function and depleting variants with low function (Fowler and Fields 2014). Functional selections can take many forms, but some of the most common are growth selection, measurements of ligand binding, and use of functional reporters such as fluorescence to measure stability and other functionality. With growth assays, high throughput sequencing of a variant

library across a time course can be used to determine relative variant enrichment or depletion, while selections with fluorescent reporters use fluorescence-activated cell sorting (FACS) and sequencing of sorted bins to determine relative variant function.

The number of MAVEs performed has grown in recent years, providing functional annotation for over 200,000 total variants (Weile and Roth 2018), and covering dozens of genes across many different organisms. Deep mutational scanning approaches have the potential to transform pharmacogenomic implementation (Chiasson et al. 2019), but have only been performed with a handful of pharmacogenes, including *TPMT*, *NUDT15*, and *VKORC1*. To measure variant abundance in a multiplexed fashion, a method using a fusion GFP reporter was developed, called variant abundance by massively parallel sequencing (VAMP-seq). This approach was used with the pharmacogene *TMPT*, which metabolizes thiopurine drugs (Matreyek at al. 2018). VAMP-seq was also used with NUDT15 and VKORC1, in combination with a deep mutational scan of VKOR activity and a deep mutational scan of NUDT15 drug sensitivity (Suiter et al. 2020, Chiasson et al. 2020). VAMP-seq has also been used to measure the function of a hundreds of *CYP2C9* and *CYP2C19* variants (Zhang et al. 2020). To date, no one has performed a comprehensive deep mutational scan on the activity of a human CYP enzyme, even though there is a clear need for such comprehensive functional annotation.

1.8 Specific Aims

1.8.1 Engineering S. cerevisiae to improve heterologous CYP activity

S. cerevisiae has been used to express and characterize human CYP enzymes for decades, and is an outstanding heterologous expression system. The most common yeast strain used for

such purposes is laboratory strain W303 with added human P450 accessory proteins cytochrome P450 reductase and cytochrome b5 (Pompon et al. 1994). Recently, several studies have shown how targeted strain modifications can also improve heterologous CYP activity in yeast (Emmerstorfer et al. 2015, Michener et al. 2012). Increasing heterologous CYP expression levels is important for multiplexed functional characterization of variants, so there is a need to characterize the best combination of yeast strain modifications for optimal human CYP expression. Furthermore, no studies have comprehensively screened yeast natural isolates for improved CYP expression, even though natural variation can be a source of a wide range of phenotypes. In this aim, I took a multi-pronged approach to improving heterologous CYP expression in yeast by testing targeted strain modifications, addition of human accessory proteins, and screening of natural isolates to optimize expression and activity of human CYP2C9, CYP2C19, and CYP2D6. These strains and modification were screened individually and in combination with the goal of generating the optimal yeast P450 expression strain.

1.8.2 Development of a pooled cell-based CYP activity assay in yeast

Existing assays of P450 activity are often performed on cell extracts *in vitro* or are only suited to measuring the activity of single P450 variants. In this aim, I developed a yeast-based CYP activity assay capable of measuring the activity of thousands of variants in a pooled format. I tested three different approaches for measuring CYP2C9 activity: fluorogenic substrates, growth inhibition assays, and activity-based protein profiling. Fluorogenic substrates are a common method of measuring CYP activity rapidly and can be performed in cells, but have not been developed for pooled assays. Substrates that lead to CYP-dependent growth inhibition have been documented for a variety of CYP enzymes, including CYP2C9, but do not have

demonstrated cytotoxicity in yeast. Finally, activity-based probes have been developed for human liver P450s for activity-based protein profiling (Wright et al. 2009), but had not yet been evaluated for labeling in yeast.

1.8.3 Massively parallel functional profiling of CYP2C9 variants

Cytochrome P450 enzymes play a central role in Phase I drug metabolism, and polymorphisms in these enzymes lead to a wide range of metabolic activity among different patients. Dosing guidelines are available for some of the more common CYP alleles, but patient genotyping often results in variants of unknown significance (VUS) which are uninformative in clinical settings. There is a need for comprehensive CYP variant functional annotation to improve PGx implementation in routine clinical care, and to reduce adverse effects for patients. I performed a deep mutational scan of CYP2C9 (a pharmacogene important in metabolizing warfarin, phenytoin, and many other drugs including a number of NSAIDs) by leveraging the yeast strain and pooled activity assay developed in the first two aims. By coupling the activity assay with fluorescence-activated cell sorting and high throughput sequencing, I generated functional measurements for thousands of CYP2C9 variants, giving new insights into CYP2C9 biochemistry and structure. I also report results from a parallel deep mutational scan of CYP2C9 abundance in human cells. The activity assay developed in these aims can be extended to test different-activity based probes as well as other CYP enzymes.

Chapter 2. Engineering *Saccharomyces cerevisiae* to improve heterologous CYP activity

---

The work described here was performed by Clara Amorosi and collaborators. Cloning of CYP vectors and western blots were conducted by Kate Sitko. Microsomal preparations were conducted by Clara Amorosi, Sutapa Ray, and Kate Sitko. CO binding spectra were conducted by Sutapa Ray and Clara Amorosi.

Saccharomyces cerevisiae has long been used as a heterologous system of mammalian CYP expression, due to ease of genetic manipulation and ability to express human CYPs without modification. However, addition of human CYP accessory proteins is necessary to reconstitute activity in most yeast expression systems. In addition to humanizing a laboratory yeast strain, we tested additional targeted strain modifications and found that certain modifications including knocking out endogenous proteases produced slight increases in human CYP activity and expression. Moreover, screening a set of 11 *S. cerevisiae* natural isolates from a variety of sources revealed a range of differences in CYP expression level and activity. We followed up on Y12, a sake strain that showed much greater human CYP activity than the lab strain, possibly due to increased expression levels. However, this strain also showed several undesirable cell aggregation-related phenotypes. This approach highlights the utility and challenge of working with natural isolates.

2.1 Introduction

Saccharomyces cerevisiae has been used as a model system for recombinant P450 expression for over 30 years (Oeda et al. 1985) and can be engineered to express highly active human P450 enzymes, with the added benefit that recombinant expression allows for the isolated

study of a single P450 isoform. Unlike humans, which have 57 CYP genes, *S. cerevisiae* has only three endogenous CYPs (Kappeli 1986): *CYP51*, *CYP56*, and *CYP61*, all of which are involved in biosynthesis rather than metabolism. *CYP51* (*ERG11*) is involved in ergosterol biosynthesis (Karst and Lacroute 1977), and *CYP61*, though putative, is thought to also play a role in ergosterol biosynthesis (Pompon et al. 1994). *CYP56* is only expressed during sporulation and is important for spore wall maturation (Bogengruber et al. 1998). Thus, when using yeast as a vector for recombinant CYP expression, background effects from native yeast CYPs should be negligible when testing metabolic CYPs. In addition to CYPs, yeast also possess a cytochrome P450 reductase (*NCP1*) and a cytochrome b5 (*AIM33*), which share 33% and 30% amino acid sequence identity with the homologous human genes, respectively. Interestingly, Ncp1 can inefficiently couple (bind and transfer electrons) with human CYPs, but Aim33 has minimal to no coupling with human CYPs (Pompon et al. 1994). Thus typical yeast expression systems often co-express heterologous cytochrome P450 reductase (CPR, human gene name *POR*) and cytochrome b5 to reconstitute full activity, "humanizing" the strain. Humanized systems expressing human genes or even entire pathways in yeast have been utilized to understand many different aspects of human biology (Laurent et al. 2016), including P450 biochemistry. Stoichiometry of CYP and accessory proteins is important, and it is estimated that in human liver membranes, CYPs outnumber CPRs 10-20:1 (Backes and Kelley 2003), but since CYP-CPR coupling occurs in a 1:1 fashion, increasing the amount of available reductase may lead to a higher level of overall CYP activity. However, human CPR is toxic at high levels to yeast (Urban et al. 1993).

Although most studies that use yeast to express mammalian CYPs use *S. cerevisiae*, several studies have expressed various CYPs in the fission yeast *Schizosaccharomyces pombe* for increased expression levels (Drăgan et al. 2011, Neunzig et al. 2013). However, fission yeast are less genetically tractable than *S. cerevisiae*. In *S. cerevisiae*, strain engineering efforts to improve CYP expression and activity have been piecemeal rather than thorough. Common methods involve expressing human CYPs in strain W303 and co-expressing human CPR and b5 (Pompon et al. 1994). Yeast strains with accessory proteins(s) can achieve a specific content (proportional to total microsomal protein) of 50-500 pmol P450 per mg of microsomal protein (Pompon et al. 1995).

Some studies have focused on gene knockouts or overexpressions to improve CYP activity. Overexpression of endoplasmic reticulum (ER) membrane protein Ice2 has been shown to stabilize CPR through changing ER morphology (Emmerstorfer et al. 2015), though the mechanism is not well understood. Additionally, overexpression of heme biosynthesis gene *HEM3* alleviates heme limitation when expressing heterologous CYPs (Michener et al. 2012), although increasing heme production too much can lead to porphyrin toxicity. Finally, protease deficient strains have been used to achieve relatively high amounts of CYP2C19 expression in yeast (Wang et al. 2011), specifically by knocking out *PEP4* and *PRB1*, two main endoproteinases in yeast (Jones 1991).

The goal of this work was to determine the best set of strain modifications for expressing highly active human CYP enzymes in *S. cerevisiae*. We specifically focused on a S288C derivative strain, due to the wide range of S288C strains and auxotrophic markers available in the Dunham lab. Many studies expressing recombinant CYPs in yeast have used W303 as a

starting strain background (Pompon et al. 1996, Peyronneau et al. 1992, Hamann and Møller 2007, Michener et al. 2012), and one of the main differences between S288C and W303 is that S288C has an inactivated *HAP1*, which encodes a transcription factor responsive to heme and oxygen levels (Gaisne et al. 1999). To determine if this is important to CYP activity, we tested a S2888C derivative where *HAP1* has been corrected to wild type (source: Mark Hickman and Fred Winston), as well as a number of different strain modifications based on the literature.

In contrast to well-behaved laboratory strains, natural isolates can be a source of a wide range of traits (Liti 2015), and in the case of S288C, the phenotypes of the lab strain are often the exception, rather than the rule (Warringer et al. 2011). In settings where targeted strain modifications do not produce the desired phenotype, screening natural variants can be an advantageous approach (Steensels et al. 2014). We screened 11 strains from the Saccharomyces Genome Resequencing Project (Liti et al. 2009), a collection of *S. cerevisiae* strains from diverse ecological and geographical niches. We determined basal CYP activity levels across these strains and chose the best strain for further targeted strain engineering. Ultimately, we found that a joint approach leveraging both natural variation and targeted engineering can produce markedly better results than performing the individual approaches separately.

2.2 Results

We expressed four different human CYP isoforms in a S288C derivative strain of *S. cerevisiae* and measured expression level differences between isoforms. We engineered the S288C derivative to express human CYP accessory proteins cytochrome P450 reductase (CPR) and cytochrome b5, leading to large increases in CYP activity. We also tested the effect of targeted gene overexpression and knockouts on CYP expression and activity. Finally, we

screened a set of *S. cerevisiae* natural isolates to determine whether any had intrinsically higher CYP expression and activity, and found that one natural isolate, Y12, had intrinsically higher levels of CYP activity compared to the lab strain.

2.2.1 Expression level differences between CYP isoforms

We focused our investigation of *CYP2C9*, *CYP2C19*, and *CYP2D6* due the high levels of polymorphisms in these genes and the clinical relevance of their drug targets (Gordon et al. 2014, Ji et al. 2016). We also evaluated expression of human CYP4Z1, a relatively newly discovered CYP (Rieger et al. 2004) that has proven difficult to express recombinantly (A. Rettie, personal communication). *CYP2C9*, *CYP2C19*, *CYP2D6*, and *CYP4Z1* were codon optimized, synthesized (IDT), and cloned into low copy plasmids under an inducible galactose promoter. CYPs were C-terminally tagged with HA in order to avoid disruption of the terminus anchored to the ER. We transformed a S288C strain with each CYP vector to measure basal specific content and expression levels of the individual isoforms. Relative CYP expression was measured in each strain by blotting for the CYP HA epitope tag and performing CO difference spectra on yeast microsomes (the cell fraction that contains the ER). In this assay, reduced CYP holoenzyme reacts with carbon monoxide (CO) and forms a complex that produces a characteristic wavelength. Measuring the CO difference spectra is the gold standard for determining the amount of CYP holoenzyme, but is a laborious procedure that limits such analysis to no more than a handful of individual strains.

We collected CO difference spectra to determine CYP activity in various strains. This spectral assay measures the level of active CYP by looking for the characteristic peak of reduced CO-bound CYP at ~450 nm. **Figure 2.1** shows the spectra of microsomes from strain YMD3289

(S288C derivative with *HAP1+*) expressing different CYPs. CYP2C9 showed a relatively "pure" spectrum, with a single peak at ~450 nm. CYP2C19, CYP4Z1, and CYP2D6 had additional peaks at ~430nm, which indicated the presence of inactive CYP or other heme proteins. The control strain had no peak at 450 nm (not shown), indicating that endogenous yeast CYPs were not present in high enough concentration to interfere with the signal from human CYPs. From this spectra, CYP specific contents were determined (**Table 2.1**), showing that the order of active CYP expression in microsomes is roughly 2D6 > 2C9 > 4Z1 > 2C19.

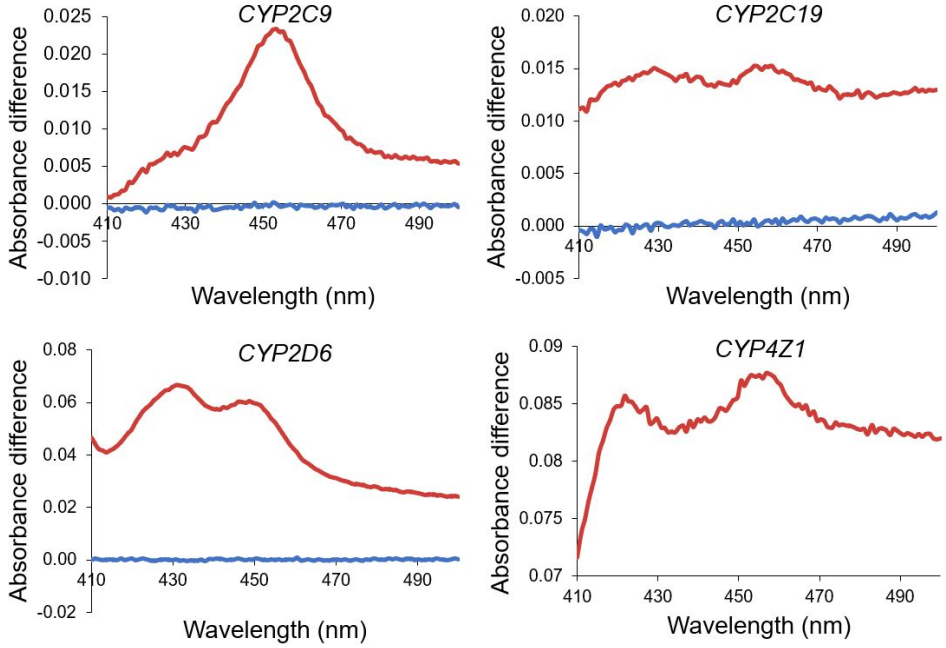

Figure 2.1. CO difference spectra of CYP isoforms expressed in yeast microsomes. CYPs expressed on a low copy plasmid in a S288C derivative strain YMD3289, harvested after 15 hours galactose induction. The blue line is the baseline reading after sodium dithionite reduction, while the red line is the CO-bound spectra.

Table 2.1. CYP specific content in yeast microsomes

CYP	P450 concentration (nM)	Total protein (mg / mL)	Specific Content (pmol P450 / mg microsomal protein)
CYP2C9	300	8.42	35.6
CYP2C19	60	7.24	8.3
CYP2D6	500	6.84	73.1
CYP4Z1	100	6.96	14.37

Microsomes isolated using enzymatic digestion and polyethylene glycol precipitation from strain YMD3289. Concentration determined from CO difference spectra (**Figure 2.1**), and microsomal protein quantified using a Qubit Protein Assay Kit (Thermo).

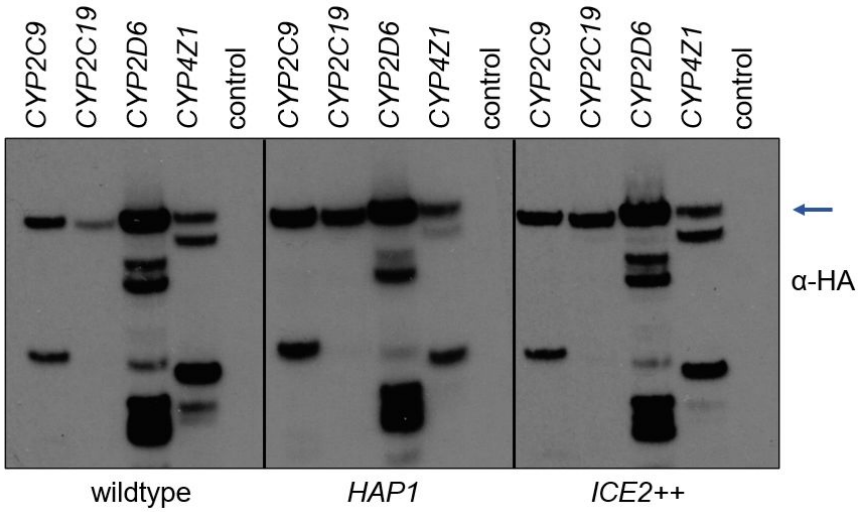

Figure 2.2. Western blots of CYPs in three S288C strains. Four different CYPs (α-HA) expressed in three different S288C backgrounds: wild type (CJA25), *HAP1+* (CJA47), and overexpressed *ICE2* (CJA57), after overnight galactose induction. Full length protein shown by blue arrow, control is strain with empty vector.

CYP expression was also validated via western blot in a wildtype S288C strain and a strain with fixed *HAP1* (normally non-functional in S288C). Relative expression levels of full-length protein followed the results from the CO difference spectra, with CYP2D6 > CYP2C9 > CYP2C19, while CYP4Z1 showed expression levels similar to CYP2C19 depending

on the strain (**Figure 2.2**). We found that CYP2D6 showed the greatest amount of degradation, and CYP2C19 the least. The presence of a functional Hap1 appeared to increase the amount of full length CYP2C9, CYP2C19, and CYP4Z1.

2.2.2 Humanization of S. cerevisiae *lab strain*

Because the endogenous yeast reductase couples inefficiently to human CYPs (Urban et al. 1994), it was necessary to express human CPR and b5 in our strain to "humanize" it. Human cytochrome P450 reductase (CPR) and cytochrome b5 sequences were codon optimized, synthesized (IDT), and cloned into low copy plasmids. CPR was C-terminally tagged with FLAG, and b5 was N-terminally tagged with MYC. Human CPR was cloned under an inducible galactose promoter to mitigate toxicity, and b5 was tested under an inducible promoter and a strong constitutive promoter *GPD* (Mumberg and Funk 1995) in order to determine the optimal CYP:CPR:b5 stoichiometry. We found that b5 worked best under a constitutive promoter rather than an inducible promoter (**Figure 2.3A**). To fully humanize the strain, both CPR and b5 were integrated into the genome, and activity was validated using a fluorogenic activity assay that measures the turnover of CYP2C9 substrate BOMCC into a fluorescent product. We found that introducing human CPR increases the rate of BOMCC turnover 5X while introducing both human CPR and b5 increased the rate 16X (**Figure 2.3B**). Adding human accessory proteins was key to expressing highly active CYP2C9.

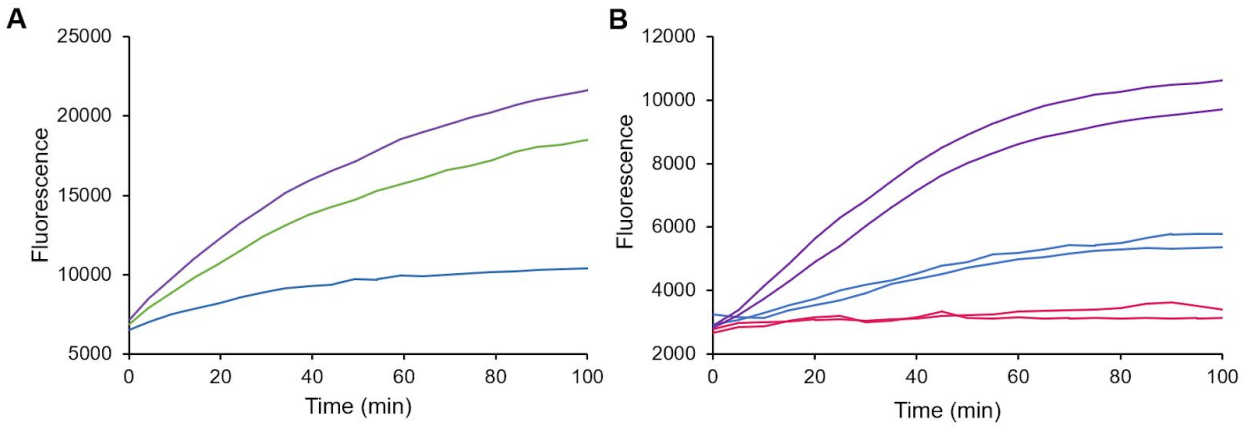

Figure 2.3. Fluorogenic assay of BOMCC turnover by CYP2C9 with and without accessory proteins. A) BOMCC turnover in strain background YMD4254 (CPR integrated strain) with empty vector in blue, YMD4254 with GAL1-b5 in green, or YMD4254 with GPD-b5 in purple. All strains expressing CYP2C9 vector. 2.5 OD cells used for each sample, fluorescence measured with 70% gain. B) BOMCC turnover in strain background YMD4253 (*HAP1* protease deficient strain) in red, YMD4254 (YMD4253 with integrated CPR) in blue, or YMD4255 (YMD4253 with integrated CPR and b5) in purple. All strains expressing CYP2C9 vector and biological replicates shown. 1 OD cells used for each sample, fluorescence measured with 60% gain. To calculate rate of turnover, average slope from 0 minutes to 60 minutes was used. For A) and B), 50 uM BOMCC added to each sample.

2.2.3 Other strain engineering targets and combinatorial testing of modifications in lab strain

After examining studies that expressed human and other heterologous CYPs in yeast, we determined several strain modifications that might increase CYP activity or expression. Using homologous recombination, strains overexpressing *HEM3* (under the constitutive *TEF* promoter) and overexpressing *ICE2* (under the inducible *GAL1* promoter) were constructed, using auxotrophic markers *HIS3* and *LEU2* respectively. Heme overexpression in *HEM3*++ strains was confirmed using the oxalic acid fluorescence assay (Michener et al. 2012) and resulted in an approximate 3-fold increase in intracellular heme levels (**Figure 2.4**). The S288C derivative lab strain was also tested with a functional *HAP1* (fixed from "wildtype" nonfunctional state) as Hap1 is important in heme regulation. All strain modifications were made in haploid cells, and

strain crosses, sporulations, and selections were performed to obtain combinations of strain modifications.

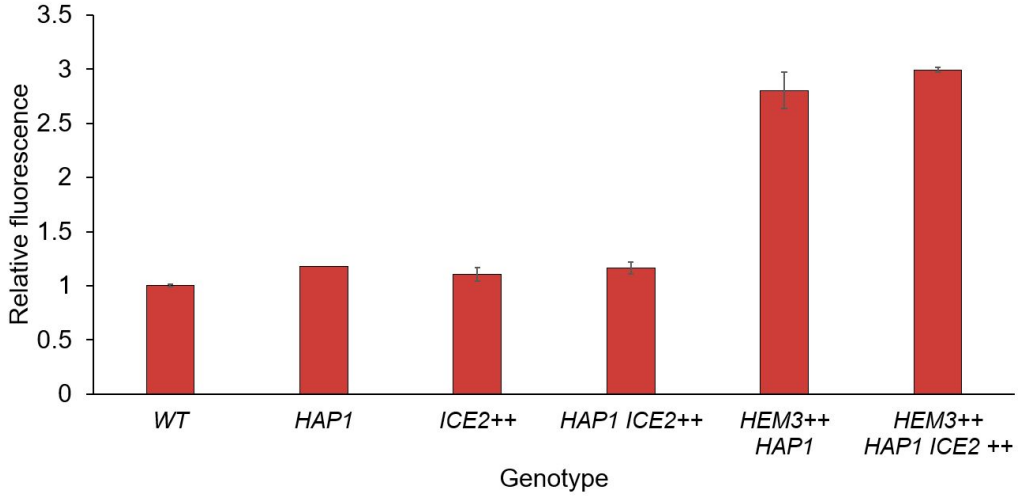

Figure 2.4. Intracellular heme content in S288C strains with targeted gene modifications. Heme content measured by the oxalic acid fluorescence assay (Michener et al. 2012). *WT* = wildtype (nonfunctional *HAP1*), *HAP1* = functional (fixed) *HAP1*, *ICE2*++ = *ICE2* overexpression, *HEM3*++ = *HEM3* overexpression. From left to right, strains are CJA71, YMD3289, CJA87, CJA81, CJA74, CJA88. Error bars show standard deviation of two replicates.

Additionally, the two main proteases (*PEP4* and *PRB1*) were knocked out via a scarless integration method (Dong et al. 2013) as a direct method of increasing CYP expression by blocking protein degradation. The effectiveness of this modification was assessed via western blot in strains expressing CYP2D6 from a low copy plasmid, induced for a short period of time. Protease knockouts resulted in the greatest increase in CYP2D6 expression, followed by having a functional *HAP1* (**Figure 2.5**). Overexpression of *ICE2* and *HEM3* appeared to have no effect on CYP2D6 expression. The effect of *HAP1* was also seen with CYP2C9 and CYP2C19 (**Figure 2.2**)

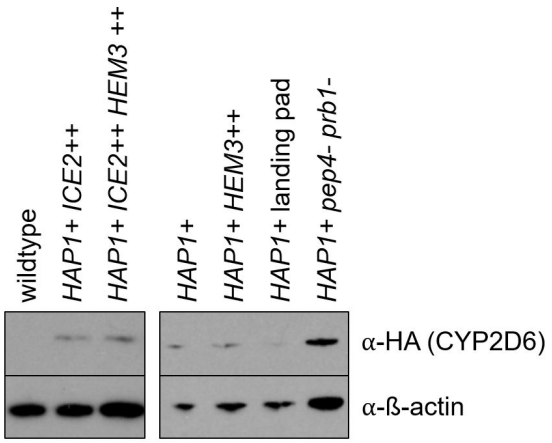

Figure 2.5. CYP2D6 expression in S288C strains with targeted gene modifications. Western blots of CYP2D6 (α-HA) and *β*-actin load control (α-*β*-actin) in various strains. Strains induced in galactose for 1hr. Strains from left to right: CJA71 (wildtype), CJA81 (*HAP1*+ *ICE2*++), CJA88 (*HAP1*+ *ICE2*++ *HEM3*++), YMD3289 (*HAP1*+), CJA74 (*HAP1*+ *HEM3*++), CJA99 (*HAP1*+ landing pad), YMD4253 (*HAP1*+ *pep4- prb1-*).

The effect of these strain modifications on CYP was assessed with the BOMCC fluorogenic assay and CYP2C9. In search of an elusive "maximum activity" strain, we crossed the humanized and protease deficient strain expressing CYP2C9 (YMD4256) to the strain where *ICE2* and *HEM3* are overexpressed (CJA88). Because each modification (except for the scarless protease knockouts) was marked with a separate auxotrophic marker, we were able to select and screen different combinations of modifications easily. We sporulated the cross and dissected tetrads, selecting for CYP2C9 using G418 drug selection. 36 segregants carrying the CYP2C9 vector also carried human CPR and b5, and of these, 9 had wildtype *ICE2* and *HEM3*, 13 had wildtype *ICE2* and overexpressed *HEM3*, 5 had overexpressed *ICE2* and wildtype *HEM3*, and 9 had both *ICE2* and *HEM3* overexpressed. We tested the turnover of BOMCC with all 36 of these strains, and found that humanized strains with overexpressed *HEM3* showed a modest increase in BOMCC turnover relative to the parent humanized strain, regardless of *ICE2* status. These 36

segregants had all possible combinations of *PEP4* and *PRB1* knockouts (both wildtype, one wildtype, both knocked out), but we were unable to determine whether protease deficiency changed BOMCC turnover rate due to low sample numbers, though we had already determined that protease deficiency greatly increased CYP2D6 expression levels (**Figure 2.5**).

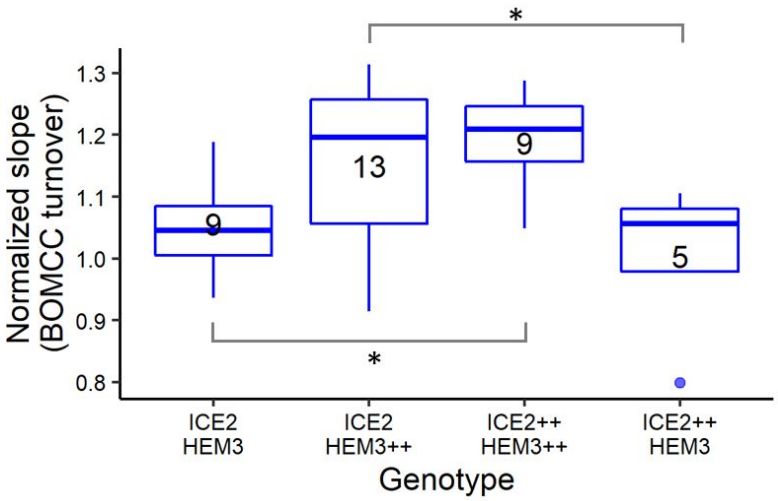

Figure 2.6. Rate of BOMCC turnover in humanized strains with *ICE2* or *HEM3* overexpression. Boxplots of BOMCC turnover rate in humanized strain background expressing CYP2C9 with the following genotypes: *ICE2 HEM3*: wildtype *ICE2* and *HEM3*, *ICE2 HEM3++:* wildtype *ICE2* and overexpressed *HEM3*, *ICE2++ HEM3++:* overexpressed *ICE* and *HEM3*, and *ICE2++ HEM3:* overexpressed *ICE2* and wildtype *HEM3*. Strains isolated from a cross between humanized strain YMD4256 and *HEM3 ICE2* overexpression strain CJA88, and the number of (haploid) segregants of each genotype is shown. BOMMC turnover rate calculated from slope of fluorescence value over time from 5 to 50 minutes, normalized to parent humanized strain. Significance calculated with a two-sided t-test, * indicates $p < 0.05$.

2.2.4 Screening yeast natural isolates for CYP activity

In addition to testing target strain modifications in a laboratory strain, we screened a set of natural isolates from different ecological and geographic niches from the Saccharomyces Genome Resequencing Project (SGRP) (Liti et al. 2009). We chose 11 strains that showed minimal flocculation phenotypes and other related aggregation traits (Hope and Dunham 2014),

as clumping phenotypes present challenges with many standard yeast experiments, and are especially an issue with flow cytometry. In fact, lack of flocculation due to a mutation in transcription factor *FLO8* is a key feature of S288C (Liu et al. 1996). As an initial test of activity, each of the 11 strains was transformed with a CYP2D6 vector under a galactose-inducible promoter, because CYP2D6 was expressed at a higher level compared to CYP2C9 and CYP2C19 in the lab strain. Strains were induced, harvested, and activity was measured by monitoring the turnover of a fluorogenic substrate, EOMCC. In general, most strains showed similar or reduced EOMCC turnover compared to the lab strain, with DBVPG6044, Y55, and YS9 showed negligible turnover altogether. However, sake strain Y12 stood out with more than twice the fluorescence increased compared to the lab strain (**Figure 2.7**).

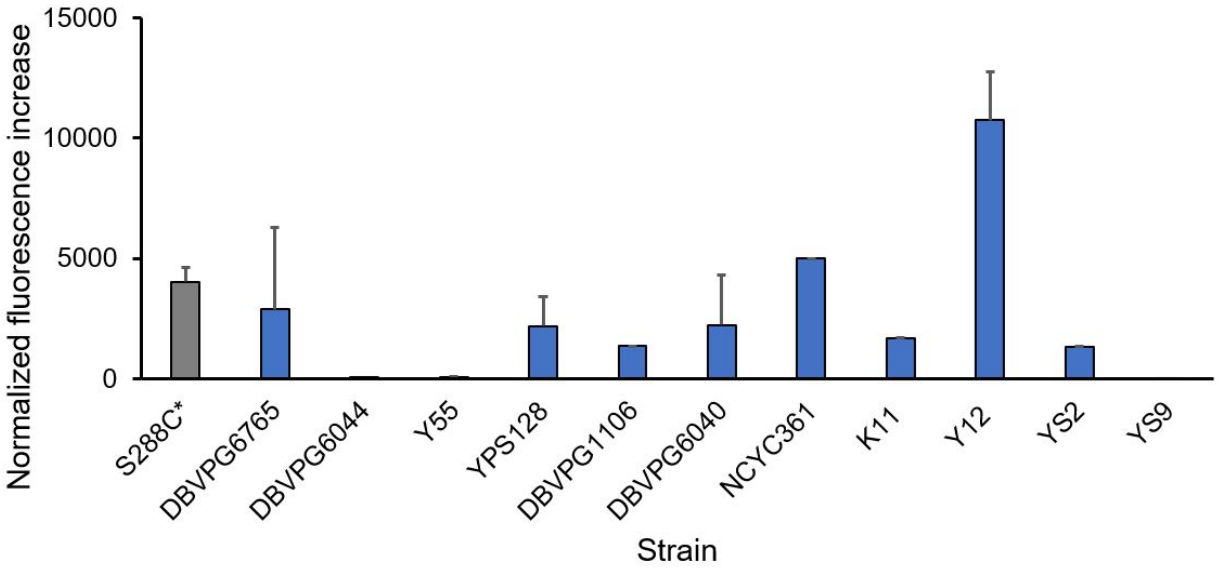

Figure 2.7. EOMCC turnover by CYP2D6 in non-flocculent natural isolates of *S. cerevisiae*. 11 strains from the SGRP collection (Liti et al. 2009) were transformed with CYP2D6 and activity was measured with the fluorogenic substrate EOMCC. Strain name listed, S288C* is YMD4253. Value shown here is the fluorescence accumulated after 4 hrs, normalized to the fluorescence of the strain background. Standard deviation of biological replicates shown when available. DBVPG6044, Y55, and YS9 showed

minimal fluorescence increase with CYP2D6. Cells harvested after 8 hrs of induction in galactose and 2.5 OD cells used for each sample, fluorescence measured with 70% gain. 50 µM EOMCC added to each sample.

As all natural isolates were screened without human accessory proteins, we followed up on Y12 by humanizing it (integrating human accessory proteins CPR and b5). A previous study had engineered a set of auxotrophic markers into Y12 and three other natural isolates (Louvel et al. 2014), and we were able to use this strain to easily humanize Y12 using selectable markers. Expression of accessory proteins was confirmed via western blot (**Figure 2.8A**) in strains expressing wildtype CYP2C9 or a null variant. Interestingly, CYP2C9 appears to be expressed at a higher level in Y12 compared to S288C. Next, we tested CYP2C9 activity in the humanized Y12 strain compared to the lab strain by measuring BOMCC turnover (**Figure 2.8B**). The fully humanized Y12 strain showed 5x BOMCC turnover compared to the protease deficient humanized lab strain (calculated from slope of fluorescence increase over time).

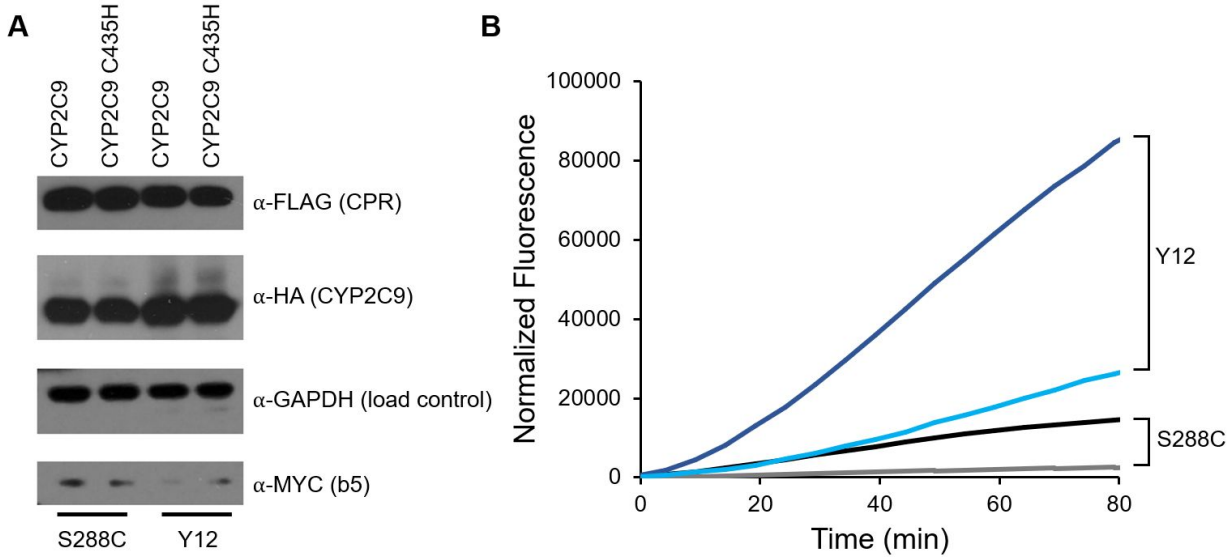

Figure 2.8. Expression level and activity of CYP2C9 in natural isolate Y12 compared to laboratory strain. A) Western blot of CPR (α-FLAG), CYP2C9 (α-HA), load control (α-GAPDH) and b5 (α-MYC) in S288C (YMD4255) and Y12 (CJA418). Strains expressed wildtype CYP2C9 and a null variant

(C435H) and induced for ~8 hrs. B) BOMCC turnover by CYP2C9 in Y12 (light and dark blue) and S288C (grey and black) humanized strain background. In dark blue, CJA418 (Y12 with integrated human CPR and b5), in light blue, CJA409 (Y12 with integrated human CPR), in black, YMD4255 (protease deficient S288C derivative with integrated human CPR and b5), and in grey, YMD4254 (protease deficient S288C derivative with integrated human CPR). 2.5 OD cells used for each sample, fluorescence measured with 70% gain, and 50 μM BOMCC added to each sample.

Though Y12 showed great promise in terms of expressing highly active CYP, we noticed that this strain exhibited several unwanted phenotypes, including slow growth while expressing human CYPs (Appendix **Figure A.1**), and showed a mild aggregation phenotype. Thus, we performed a cross between the CYP2C9-expressing humanized Y12 and humanized S288C strains in the hope that the segregants of this cross would retain the high-activity/expression phenotype of Y12 and the non-clumpy phenotype of S288C. Of more than 80 segregants typed, 35% showed evidence of a mother-daughter separation defect, 45% were flocculent, and 20% had no obvious clumping phenotype (segregants screened with microscopy). 18 of the segregants with no obvious clumping phenotype were screened for CYP2C9 activity by measuring BOMCC turnover. We found that all 18 non-clumpy segregants had BOMCC turnover levels similar to the parent S288C strain, rather than the Y12 strain. The high-activity Y12 phenotype does not appear to be linked to cell aggregation traits, as four segregants with mother-daughter separation defects and four flocculent segregants also showed S288C-like levels of BOMCC turnover (data not shown). Thus, it is likely the increased activity phenotype of Y12 is a complex genetic trait.

We next used an experimental evolution approach to select against the undesirable aggregation phenotype. We performed a modified form of a settling assay on the Y12 strain (Ratcliff et al. 2012) to select against cell-aggregation phenotypes, an anti-settling assay. Briefly, after overnight growth in glucose media, cells were allowed to settle, and non-settled cells were

taken from the top of the media used to inoculate new media. After 14 transfers (two weeks), we saw a reduction in cell aggregation while retaining CYP activity. However, we noticed that while cells no longer displayed clumping phenotypes in glucose media, they exhibited a galactose-induced flocculation phenotype. Though our experimental aggregation approach was successful in one experimental condition, Y12 showed unexpectedly complicated aggregation-related traits, making it a challenge to work with.

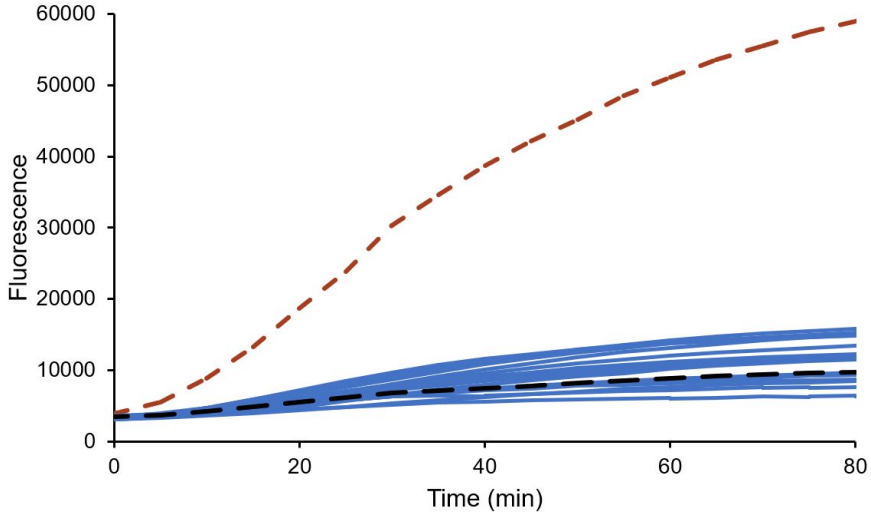

Figure 2.9. Non-clumpy segregants of Y12 x S288C cross have activity levels similar to S288C. BOMCC turnover (measured in fluorescence) by CYP2C9 in parental humanized Y12 strain CJA418 (dashed red line) and parental humanized S288C strain YMD4255 (dashed black line). In blue lines, BOMCC turnover from 18 segregants of the Y12 x S288C cross, 1 OD cells used for each sample, fluorescence measured with 60% gain, and 50 μM BOMCC added to each sample, performed in triplicate.

2.3 Discussion

Expressing human CYPs in recombinant yeast invariably requires strain engineering to improve yield or activity. We characterized human CYP expression and activity in a S288C derivative strain of *S. cerevisiae*. We were able to recapitulate known human CYP isoform-dependent expression differences in *S. cerevisiae*, measuring specific contents slightly

lower than literature values of 100 pmol/mg for 2D6, and 50 pmol/mg for 2C9, respectively (Schmid and Urlacher 2007). Degradation products of each isoform varied widely in size and quantity and did not appear to be related to strength of CYP expression, though the large proportions of degradation products observed may be due to the long induction times used. Humanizing yeast by co-expressing human accessory proteins CPR and b5 resulted in large increases in activity, as expected.

Furthermore, we attempted several targeted strain modifications aimed at increasing human CYP expression or activity. We found that having a functional *HAP1* increased expression of all CYPs tested. Though the direct mechanism behind this is not known, Hap1 is a heme-dependent transcriptional activator (Hickman and Winston, 2007) and W303, a lab strain commonly used to express human CYPs, has a functional *HAP1*. We used this *HAP1*+ strain to express and characterize CYP4Z1, which is the first time this CYP, the least well characterized CYP4 enzyme, has been expressed in a functionally useful yeast system (McDonald et al. 2017). Additionally, knocking out the two main vacuolar proteases, *PEP4* and *PRB1*, increased CYP2D6 expression. We expect this is also true for other CYPs, but did not test the effect of these protease knockouts on CYP2C9 or CYP2C19 expression.

We also tested the overexpression of heme biosynthesis gene *HEM3*, and demonstrated that it led to a 3X increase in intracellular heme levels (bound and unbound heme). We also tested the overexpression of *ICE2* as it has been implicated in CPR stabilization through ER morphology (Emmerstorfer et al. 2015). Both *HEM3* and *ICE2* overexpression did not appear to change CYP2D6 expression levels, but *HEM3* overexpression resulted in a slight (~20%)

increase in CYP2C9 activity, as measured by BOMCC turnover, which may indicate that heme is a limiting factor. *ICE2* overexpression did not appear to have an effect on CYP2C9 activity.

Overall, we performed a non-exhaustive investigation into the combinatorial effects of a set of S288C strain modifications to see how they increased human CYP activity and expression. This is complicated by the fact that the degree to which each modification affects human CYP expression or activity depends on the CYP isoform. Furthermore, stoichiometry is key in a system such as this, where CYP and CPR interact in a 1:1 fashion, and we were only able to test a handful of different promoters when generating the humanized strain. Many of our strain modification elements were expressed from a *GAL1* promoter, and perhaps led to diminishing returns when introducing many genes under control of the same promoter. This might explain why a constitutive promoter worked better for expressing human b5, and why we didn't see a strong effect from *ICE2* overexpression.

S. cerevisiae is a very genetically tractable model organism and we were able to generate a number of combinations of strain modifications easily, but in the future we would like to characterize expression and activity for all possible combinations of a set of yeast strain modifications with a single CYP isoform to better determine each CYP-optimal strain. It may be that each CYP requires a slightly different set of strain modifications for optimal activity. For example, cytochrome b5 is known to enhance catalytic activities of many human liver CYP isoforms, but not all (Porter 2002, Yamazakit et al. 2002).

As a complementary approach to targeted strain modification, we screened a set of *S. cerevisiae* natural isolates for human CYP activity and found that one strain in particular, Y12, showed intrinsically higher levels of CYP activity compared to S288C. Y12 is a Japanese sake

strain, although it was originally erroneously reported as a Ivory Coast Palm wine strain (Bergström et al. 2014). Though the S288C humanized strain had higher CYP2C9 activity levels than Y12 expressing CYP2C9 alone, the humanized Y12 strain had much higher activity than S288C humanized strain, making it an attractive option as a human CYP expression system.

However, Y12 exhibited several phenotypes that were not desirable: CYP-dependent growth inhibition (much greater than in the lab strain), and flocculation as well as mother-daughter separation defects. In particular, the cell aggregation phenotypes were especially undesirable since our goal was to perform flow cytometry on these strains. Thus, we performed a cross of Y12 with the laboratory strain, but found that haploid segregants of this cross only retained lab strain level CYP activities, showing that the increased activity phenotype seen in Y12 is likely a complex trait. Y12 has a number of missense mutations in genes related to ergosterol biosynthesis, protein degradation and other key processes, such as *ERG11*, *AIM33*, *PEP4*, *PRB1*, *ICE2*, and *HAP1* (accessed from: http://www.yeastrc.org/g2p/home.do) that may have an effect on CYP expression and/or activity. It is unclear whether the increased CYP activity shown by Y12 is a result of solely increased CYP expression or a combination of factors. Dissecting the genetic basis of Y12's increased CYP activity will require more complicated QTL mapping approaches (Wilkening et al. 2014).

However, an alternate approach to fix an undesirable phenotype is to select against it using experimental evolution. We performed an anti-settling selection on Y12 in glucose media, modified from a settling assay previously developed (Ratcliff et al. 2012). After selection, we saw a significant reduction in Y12 cell aggregation in glucose while retaining CYP activity, but discovered that Y12 additionally exhibited an unexpected galactose-induced flocculation

phenotype. Evidence for a galactose-specific flocculin has been shown in *S. pombe* (Matsuzawa et al. 2011). In the future, we can perform this same anti-settling assay in galactose media in the hopes of removing all cell-aggregation phenotypes from Y12. Coupling this experimental evolution approach with genome sequencing will allow us to map the genetic basis of Y12's many cell aggregation phenotypes. In conclusion, working with natural isolates presents unique advantages and challenges when attempting to select for or engineer a phenotype of interest.

2.4 Materials and Methods

Growth media and culturing techniques: *E. coli* were cultured at 37°C in Luria broth. Yeast were cultured at 30 °C and all yeast transformations were performed using the LiAc protocol described in Gietz and Schiestl 2007. Yeast cells carrying a CYP or variant plasmid were induced as follows: a single colony was inoculated into 5mL YPD media supplemented with 200 μg/ml G418 and grown overnight with rotation. This culture was diluted 1:50 into fresh YP media containing 2% (w/v) raffinose and supplemented with 200 μg/ml G418 and grown for at least two cell doublings. Cultures were then inoculated to OD 0.0125 into fresh YP media containing 2% (w/v) galactose and 200 μg/ml G418 and collected after 7 doublings the following day, unless otherwise specified.

Strain engineering: Genotype information for all strains used in this chapter listed in **Table 2.2**. To clone the CYP plasmids, the low-copy pRS41KGAL1 vector was constructed from the pRS416GAL1 vector (Mumberg and Funk 1995) and the pUG6 vector (Güldener et al. 1996) using Gibson assembly (Gibson et al. 2009) to clone *KanMX* into pRS416GAL1. *S. cerevisiae* codon-optimized *CYP2C9* (Uniprot: P11712), *CYP2C19* (Uniprot: P33261), *CYP2D6* (Uniprot: P10635), and *CYP4Z1* (Uniprot: Q86W10) sequences were synthesized (Integrated

DNA Technologies) with a C-terminal HA tag (sequence: YPYDVPDYA) and cloned into pRS41KGAL1 using Gibson assembly (Gibson et al. 2009). Yeast strains were transformed with the appropriate CYP vectors and transformants were selected on YPD media supplemented with 200 µg/mL G418 to maintain the plasmid.

To engineer the humanized strains, *S. cerevisiae* codon-optimized *POR* sequence (Uniprot: P16435) was synthesized (Integrated DNA Technologies) with a C-terminal FLAG tag (sequence: DYKDDDDK) and cloned into a low-copy pRS416GAL1 vector (Mumberg and Funk 1995), resulting in the plasmid pRS416GAL1-*hCPR-FLAG*. The auxotrophic marker *TRP1* was amplified from pRS414 (Mumberg and Funk 1995) and cloned into this vector, and the fragment containing both *GAL1-hCPR-FLAG* and*TRP1* was amplified with HO flanking homology, digested with DpnI (NEB R0176), and used to transform the strain YMD4253. Transformants were selected for growth on synthetic media lacking tryptophan. Finally, *S. cerevisiae* codon-optimized cytochrome *b5* sequence (Uniprot: P00167) was synthesized (Integrated DNA Technologies) with a N-terminal MYC tag (sequence: EQKLISEEDL) and cloned into a low-copy pRS416GPD vector (Mumberg and Funk 1995). A portion of the vector containing both *GPD-MYC-hb5* and *URA3* was amplified via PCR with *URA3* flanking homology, digested with DpnI, and used to transform the yeast strain by selecting for growth on synthetic media lacking uracil. This fully humanized strain was backcrossed twice with YMD4252, resulting in the strain YMD4256.

To humanize the Y12 strain, the same approach was used except with a different auxotrophic maker to select for the CPR integration. The auxotrophic marker *LEU2* was amplified from pRS414 (Mumberg and Funk 1995), cloned using restriction digest into the

vector pRS416GAL1-*hCPR-FLAG* in place of the *URA3* gene, and the fragment containing both *GAL1-hCPR-FLAG* and *LEU2* was amplified with HO flanking homology, digested with DpnI, and used to transform the the Y12 strain CJA295 by selecting for growth on synthetic media lacking leucine. Integration of b5 into this strain was performed as described above.

A previously generated S288C derivative strain YMD3289 (*MATα HAP1+ ura3Δ0 leu2Δ1 his3Δ1 trp1Δ63*) (McDonald et al. 2017) was further modified as described below. To reduce protein degradation, the vacuolar proteases *PEP4* and *PRB1* were sequentially knocked out using the pop-in pop-out method (Dong et al. 2013) using the vector pRS406 (Mumberg and Funk 1995) with flanking sequences cloned in.

To overexpress *HEM3*, wildtype *HEM3* was amplified and cloned into pRS413TEF (Mumberg and Funk 1995) resulting in pRS413TEF-*HEM3.* Next, the fragment containing *TEF1-HEM3-CYC1* (including CYC1 terminator sequence) was cloned into integrating plasmid pRS403 using restriction digest, and the resulting plasmid was linearized and used to transform strains of interest. Transformants were selected for growth on synthetic media lacking histidine. To overexpress *ICE2*, the *GAL1* promoter was cloned into pIL07 (Liachko et al. 2010), and the fragment containing *LEU2* and *GAL1* was amplified with homology to *ICE2* and its upstream region. Strains were transformed with this fragment and selected for growth on synthetic media lacking leucine.

Table 2.2. List of yeast strains used in this study

Strain Number	**Strain description**	**Genotype**	**Strain background**
CJA25	Base strain	*MATa ura3- leu2- his3-*	FY/BY
CJA47	HAP1	*MATa HAP1+ ura3- leu2- his3-*	FY/BY

CJA57	ICE2 overexpression	*MATa Pice2::LEU2-Pgal1 ura3-*	FY/BY
YMD3289 (CJA68)	HAP1	*MATx ura3Δ0 leu2Δ1 his3Δ1 trp1Δ63 HAP1+*	FY/BY
CJA71	Base strain	*MATa ura3Δ0 leu2- his3Δ1 trp1Δ63*	FY/BY
CJA74	HEM3 overexpression	*MATx ura3Δ0 leu2- trp1- HAP1+ his3::HIS3-Ptef1-HEM3*	FY/BY
CJA81	HAP1, ICE2 overexpression	*MATa HAP1+ Pice2::LEU2-Pgal1 ura3- his3- trp1-*	FY/BY
CJA87	ICE2 overexpression	*MATa Pice2::LEU2-Pgal1 ura3- his3- trp1-*	FY/BY
CJA88	HAP1, ICE2 overexpression, HEM3 overexpression	*MATa HAP1+ Pice2::LEU2-Pgal1 his3::HIS3-Ptef1-HEM3 ura3- trp1-*	FY/BY
CJA99	HAP1, landing pad	*MATa HAP1+ ura3-ybr209w::GalCre-KanMX-1/2URA3-lox66 leu2- his3- trp1-*	FY/BY
YMD4253 (CJA149)	HAP1, protease deficient	*MATx HAP1+ ura3Δ0 leu2Δ1 his3Δ1 trp1Δ63 pep4Δ0 prb1Δ0*	FY/BY
CJA295	Y12 base strain	*MATx ura3Δ0 leu2Δ0 met15Δ0 lys2Δ0*	Y12
YMD4254 (CJA379)	hCPR, HAP1, protease deficient	*MATx ura3Δ0 leu2Δ1 his3Δ1 trp1Δ63 HAP1+ pep4Δ0 prb1Δ0 HO::Pgal1-CPR-FLAG-TRP1*	FY/BY
CJA409	hCPR	*MATx ura3Δ0 leu2Δ0 met15Δ0 lys2Δ0 HO::Pgal1-CPR-FLAG-LEU2*	Y12
YMD4255 (CJA414)	hCPR, hb5, HAP1, protease deficient	*MATx ura3Δ0::Pgpd-MYC-b5-URA3 leu2Δ1 his3Δ1 trp1Δ63 HAP1+ pep4Δ0 prb1Δ0 HO::Pgal1-CPR-FLAG-TRP1*	FY/BY
CJA418	hCPR, hb5	*MATx ura3Δ0::Pgpd-MYC-b5-URA3 leu2Δ0 met15Δ0 lys2Δ0 HO::Pgal1-CPR-FLAG-LEU2*	Y12
YMD4256 (CJA606)	hCPR, hb5, HAP1, protease deficient	*MATx ura3Δ0::Pgpd-MYC-b5-URA3 leu2-his3Δ1 trp1Δ63 HAP1+ pep4Δ0 prb1Δ0 HO::Pgal1-CPR-FLAG-TRP1*	FY/BY

Strain genotypes listed, FY/BY are S288C derivative strains.

Yeast microsome protocols: Extracting microsomes (the cell fraction containing the endoplasmic reticulum, and thus the CYP proteins) from cells is a classic way to test for CYP activity, since it eliminates confounding proteins from other cell fractions. However, developing a working yeast microsome protocol has posed an unforeseen challenge, and much time was spent testing and optimizing protocols. We found that the cell breaking technique is key to the success of the protocol. We tested sonication (van der Hoeven 1981), bead beating (Pompon et al. 1996), enzymatic digestion (Pompon et al. 1996), and French press cell breaking methods. After cell breaking, microsomes can be isolated by either differential centrifugation (Stansfield and Kelly 1996) or precipitation with polyethylene glycol (Pompon et al. 1996). We found that enzymatic digestion followed by polyethylene glycol precipitation (or centrifugation) is optimal for preserving CYP activity.

Western blots: Cells were induced as described above, and 2.0 to 4.0 OD of cells was harvested for each sample and supernatant removed. 200 µL SUMEB buffer (1% SDS, 8 M Urea, 10 mM MOPS, pH 6.8, 10 mM EDTA, 0.01% bromophenol blue), protease inhibitors, and 5% BME was added to each sample, followed by 100 µL of 0.5 mm acid washed glass beads. Samples were vortexed on the highest setting for 3 min, and then incubated at 65ºC for 10 min. The lysate was transferred to a new tube, and spun at 5 minutes at max speed at 4ºC to clarify. Supernatant was transferred to a new tube and protein content quantified by Qubit Protein Assay Kit. Samples were normalized by protein concentration and run on a 4-12% NuPage Bis-Tris gel (Thermo Fisher) in 1x MOPS Running Buffer (Thermo Fisher) at 150V for ~1.5 hours. Gel was transferred to a nitrocellulose membrane using 1X NuPage transfer (Thermo Fisher) buffer with 20% methanol for 1 hour at 24V at 4ºC. The membrane was rinsed in 1x TBS-T (0.1% Tween)

for 5 minutes 5 times, blocked with 5% milk in TBS-T for 30 minutes, then rinsed in 1x TBS-T for 5 minutes 5 times. The blot was then cut according to target protein size, and primary antibody was added in 5% milk in 1x TBS-T and incubated at overnight 4ºC. For non-direct conjugate antibodies, the membrane was then rinsed in 1x TBS-T for 5 minutes 5 times and incubated with secondary anti-mouse-HRP 670 (GE Healthcare NA931V) in 5% milk in 1x TBS-T for 1 hr. Membrane was rinsed in 1x TBS-T for 5 minutes 5 times, and exposed using Thermo Fisher SuperSignal West Dura Extended Duration Substrate (Thermo Fisher) at a 1:1 ratio at room temp for 5 min, and then imaged. Antibody dilutions: anti-HA dHRP: 1:20,000, anti-GAPDH dHRP: 1:1,000, anti-beta-actin dHRP: 1:1,000, anti-FLAG: 1:3,000, anti-MYC: 1:10,000.

Fluorogenic plate reader assay in whole cells: Yeast cells carrying a CYP vector were induced as described above. After harvesting, cells were rinsed twice with 0.5 mL 50 mM KPi buffer, pH 8, and 1 OD of cells (unless otherwise specified) was resuspended in buffer and mixed together with 7-ethoxymethyloxy-3-cyanocoumarin (EOMCC) or 7-Benzyloxymethyloxy-3-cyanocoumarin (BOMCC) (50 μM) and 200 μM NADPH in 150 μL final incubation volume. Each sample was done in parallel with a no substrate control. Sample fluorescence (excitation: 410nm, emission: 460nm, gain: 60% unless otherwise specified) was recorded every 5 minutes on a BioTek Synergy H1 microplate reader at 37°C for 200 min with shaking. To determine relative activity, the fluorescence from each sample was normalized by subtracting the no substrate control. EOMCC was used to test CYP2D6 activity, while BOMCC was used to assess CYP2C9 activity.

Chapter 3. Development of a pooled cell-based CYP activity assay in yeast

---

The work described here was performed by Clara Amorosi and collaborators. Probe synthesis and IC50 shift experiments were conducted by Matt McDonald. Elisa Wong assisted early efforts at optimizing the activity-based protein profiling assay.

We tested several approaches for measuring CYP2C9 activity in yeast cells including fluorogenic substrates, growth inhibition assays, and activity-based protein profiling (ABPP) with a fluorescent reporter. ABPP was the only approach that was capable of labeling a pooled population of variants in an activity-dependent manner. To improve labeling, a set of activity-based probes designed from CYP2C9 inhibitors and substrates was modified with click "handles" for copper-catalyzed azide-alkyne cycloaddition. This click chemistry approach allows fluorescent labeling of CYP2C9 variants in an activity-dependent manner. Activity-based probes derived from tienilic acid produced the best results and were able to discriminate between wildtype, intermediate activity, and null CY2C9 alleles. A tolbutamide-derived probe had less labeling intensity and only discriminated between WT and null alleles. Designing clickable activity-based probes for a specific CYP enzyme is challenging, and we found that warfarin and cyclocoumarol-based probes exhibited little to no labeling of CYP2C9.

3.1 Introduction

When novel human P450 alleles are discovered via sequencing, the first step is to characterize the allele function by measuring activity with one or more substrates. The ability to measure the activity of hundreds or thousands of alleles at once requires the development of new techniques in pooled cell formats. Measuring P450 activity in cells can be challenging, and

standard assays of P450 activity are often conducted on microsomal preparations rather than whole cells. However, certain types of P450s assays such as fluorometric assays with fluorogenic substrates can be measured in living cells (Donato and Gómez-Lechón 2013). While this approach directly measures P450 activity, indirect methods of measuring P450 activity in cells can be performed with growth inhibition assays. Activity-based protein profiling methods are typically used to quantify relative enzyme levels, but have recently been adapted to measure CYP activity (Cravatt et al. 2008), albeit mostly in microsomal preparations. Here, we investigate ways of measuring P450 activity in pooled populations of yeast cells.

Fluorogenic substrates provide a highly sensitive readout of CYP activity and are preferable over growth based assays because the readout is more rapid and direct. There are three main classes of fluorescent compounds used for P450 catalytic assays: coumarins, fluoresceins, and resorufins (Grimm et al. 2013). Fluorogenic substrates have been identified for human liver CYPs including CYP2C9, CYP2C19, and CYP2D6 (Donato et al. 2004, Foti et al. 2010). With these substrates, there is some overlap between isoforms. For example, EOMCC is metabolized by CYP2D6 and CYP2C19, while BOMCC is processed by CYP2C9 and CYP3A4 (Ung et al. 2018). Other fluorogenic substrates are isoform-specific, for example AMMC is *O*-demethylated to AMHC by CYP2D6 and can be used as a CYP2D6-specific substrate in human liver microsomes (Chauret et al. 2010). As a general principle, fluorogenic probes work well in recombinant systems expressing a single CYP isoform.

Growth inhibition assays are an indirect method of determining CYP activity, where a substrate is metabolized by a P450 into a reactive species that is cytotoxic to the host cell. Growth inhibition assays are specific to both the expression system and P450 isoform. This has

been demonstrated with CYP4B1 and 4-ipomeanol in HepG2 cells (Wiek et al 2015), and with CYP2C9 and cyclophosphamide in human prostate tumor cells (Zhou et al. 2000). A larger screen identified CYP2C9-mediated cytotoxicity from losartan, benzbromarone, and tienilic acid in HepG2 cells (Iwamura et al. 2011), though metabolic activation by CYP2C9 of tienilic acid (López-Garcia and Mansuy 1994) and benzbromarone (McDonald and Rettie 2007) were previously known. CYP-mediated cytotoxicity has been less studied in recombinant yeast systems, though this has been shown with aflatoxin B1 and mouse CYP2A enzymes (Pelkonen et al.1994) and more recently with diclofenac and a bacterial P450 (van Leeuwen et al. 2011).

A third method of measuring CYP activity is activity-based protein profiling (ABPP). This chemical proteomic strategy covalently labels proteins using an active-site directed small-molecule probe, followed by click chemistry (commonly copper-catalyzed azide-alkyne cycloaddition, or CuAAC) to link a reporter molecule of interest. This allows the targeted labeling of an enzyme or enzymes of interest, and is commonly used for functional proteomics (Cravatt 2008). Recently, a set of CYP-selective covalent small molecule inhibitors (activity-based probes) have been developed for measuring CYP activity using ABPP (Wright and Cravatt 2007, Wright et al. 2009). These probes were developed by modifying existing covalent inhibitors by addition of a click chemistry “handle”: an acetylene moiety for oxidation to an electrophilic ketene that binds covalently to the enzyme. 2-ethynylnaphthalene (2EN) is a well-studied mechanism-based inhibitor of CYP2B1 and others (Roberts et al. 1993, Kent et al. 2001). 2EN was successfully modified into the clickable probe 2EN-ABP, which is capable of labeling a number of different P450 isoforms (Wright and Cravatt 2007). CYP-specific ABPP

probes have generally been used for *in vitro* assays (commonly CYP-rich microsomal preparations), rather than cell-based assays.

In this chapter, we discuss various approaches taken to develop a high-throughput pooled activity assay for CYP2C9, chosen because of its clinical relevance. This is a challenging goal and requires the assay meet the following criteria: 1) that the assay can be performed in whole yeast cells, 2) that the activity assay can differentiate between WT and inactive alleles (either through a fluorescent reporter or indirectly through growth differences), and 3) that the assay works in a pooled population of cells expressing different variant alleles (i.e., it is non-leaky). In particular, CYP activity assays already exist that fulfill the first two criteria, but bystander effects in terms of metabolite diffusion between cells is a known problem with pooled assays. We were unable to solve this problem with fluorogenic substrates, where the fluorescent product readily diffused into media and other cells in a pooled population. Instead, we turned to the covalent labeling method ABPP to develop a pooled activity assay for CYP2C9 in yeast cells. This approach is also amenable to other CYPs.

3.2 Results

We tested fluorogenic substrates, growth inhibition assays, and activity-based protein profiling in yeast cells expressing CYP2C9 to develop a pooled activity assay. Fluorogenic substrates were a rapid and accurate method of measuring CYP activity for individual alleles, but the fluorescent metabolite 3-cyano-7-hydroxycoumarin (CHC) readily traveled between cells and was not amenable to a pooled approach. Next, we tested three available CYP2C9 substrates for growth inhibition: benzbromarone, amiodarone, diclofenac, but found that none resulted in CYP-dependent cytotoxicity in yeast. Finally, covalent labeling in an activity-dependent manner

using ABPP was successful as a pooled activity measurement, but required the development of new CYP2C9 activity-based probes. We tested a set of CYP2C9 activity-based probes and found that a tienilic-acid probe resulted in the best labeling in yeast cells.

3.2.1 Evaluating fluorogenic probes in whole cell mixing assays

Fluorogenic substrates have been developed for all the major human liver P450 enzymes and are commercially available (Trubetskoy et al. 2005). CYP2C9 fluorogenic substrates include 7-methoxy-4-(trifluoromethyl)-coumarin (7-MFC), dibenzylfluorescein (DBF), 7-benzyloxy-methyloxy-3-cyanocoumarin (BOMCC), N-octyloxymethyl-resorufin (OOMR), and benzyloxy-methyl-fluorescein (BOMF) (Foti et al. 2010). Coumarins have excellent fluorescent properties (Grimm et al. 2013) and so we focused on BOMCC, which is 7-dealkylated by CYP2C9 into the fluorescent product CHC (**Figure 3.1A**).

We tested BOMCC metabolism by CYP2C9 with several alleles of known metabolic effect. We generated CYP2C9 variants *2 (R144C) and *3 (I359L), which show 60% and 20% of wildtype activity levels *in vivo*, respectively (Vogl et al. 2015**,** Lee et al. 2003, Scordo et al. 2002, Caraco et al. 2001), though levels can vary depending on substrate. These alleles are common: *2 has a minor allele frequency (MAF) of 11.7% in Europeans and 6.6% in admixed Americans, while *3 has a MAF of 11.3% in South Asians and a MAF of 5.6% in Europeans (Zhou et al. 2017). We also tested a catalytically inactive variant, C435H, where the heme-coordinating cysteine is mutated such that it is expected to have no activity. These variants were expressed in a protease deficient humanized yeast strain YMD4256 (see Chapter 2 for strain engineering details) and subjected to the BOMCC assay (**Figure 3.1B**). We found that the

relative rate of BOMCC turnover was 44% wildtype for *2, 29% wildtype for *3, and <10% wiltype for C435H.

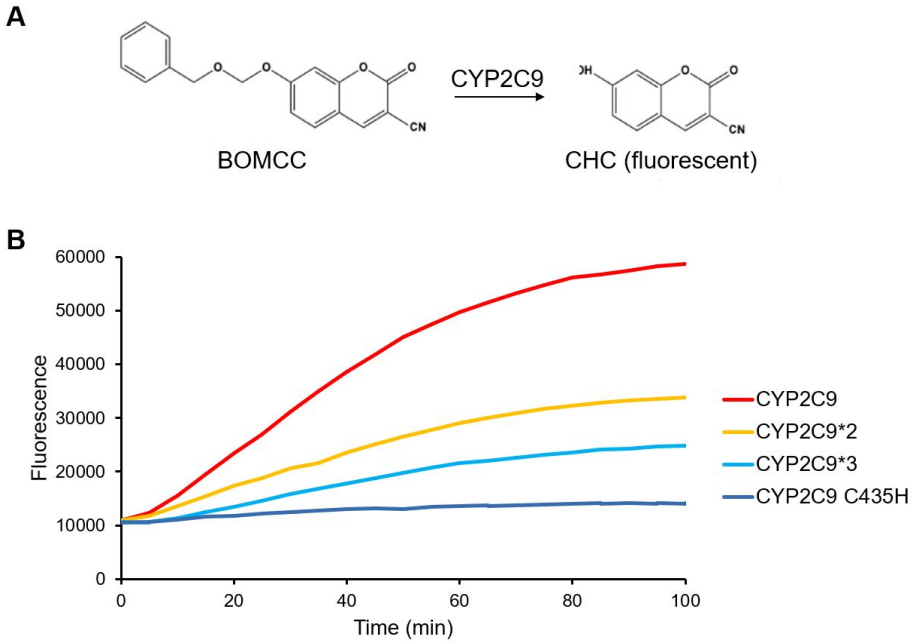

Figure 3.1. Fluorogenic assay of BOMCC turnover by CYP2C9 alleles. A) CYP2C9 metabolizes 7-benzyloxy-methyloxy-3-cyanocoumarin (BOMCC) into 3-cyano-7-hydroxycoumarin (CHC). B) BOMCC turnover in strain background YMD4256 expressing CYP2C9 wildtype (red), CYP2C9*2 (yellow), CYP2C9*3 (light blue), or CYP2C9 C435H (dark blue) plasmid. 50 µM BOMCC and 1 OD cells used for each sample, fluorescence measured with 70% gain. To calculate rate of turnover, slope from 5 minutes to 50 minutes was used.

Next, we determined whether the fluorescent product would diffuse between cells in a mixed population of alleles. We determined that during incubation with a fluorogenic probe, fluorescence readily diffused (or was exported) into the supernatant (this was tested with EOMCC, rather than BOMCC, but fluorescent product is the same) (**Figure 3.2A**). This was true even in the natural isolate strain Y12. Additionally, yeast cells are permeable to the fluorescent product (CHC), so in a mixed allele population individual cells show intermediate fluorescent levels, rather than the level relative to the variant expressed in that cell (**Figure 3.2B**). Ideally,

we would want to see two separate populations in a 1:1 mix of WT:null cells. The diffusion of CHC is perhaps unsurprising given that yeast cells are permeable to fluorogenic substrates, and because of this we turned to alternate methods of measuring CYP2C9 activity.

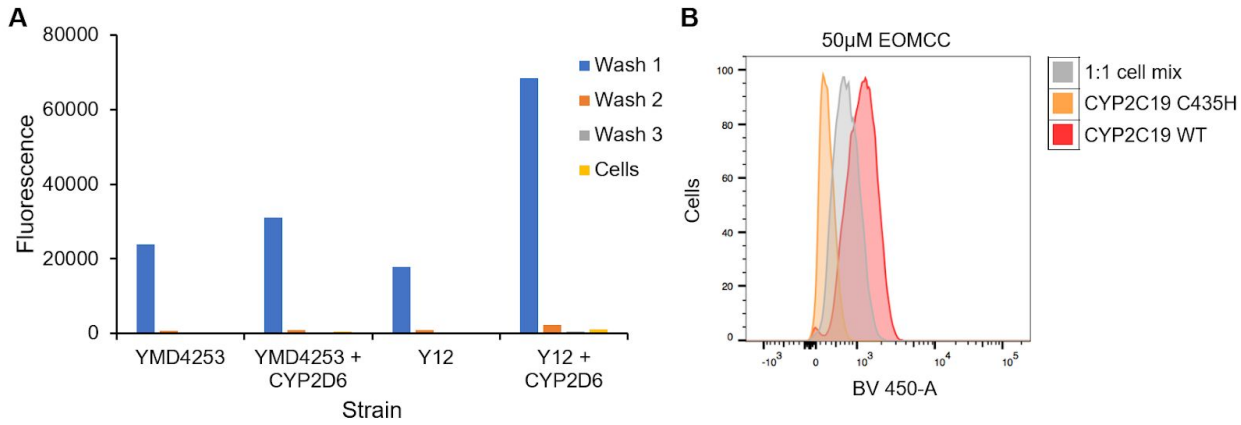

Figure 3.2. Diffusion of fluorescent product CHC. A) Fluorescence endpoint assay of yeast cells after incubation with 50 μM EOMCC overnight. Wash refers to supernatant fluorescence reading after washing cells with KPi buffer, measured after first, second and third washes. Cells refers to remaining fluorescence measurement in cells after washes. CYP2D6 expressed in YMD4253 (protease deficient S288C derivative) or Y12 (natural isolate strain), and 7.5 OD cells used for each sample. B) Flow cytometry of cells incubated with 50 μM EOMCC overnight and excess fluorescence removed via washing. CYP2C19 or null allele C435H expressed in YMD4253 (10 hour induction), and incubated with EOMCC individually or in a mixed population (1:1 cell mix). ~20,000 events recorded for each sample.

3.2.2 Testing for CYP2C9-dependent growth inhibition in yeast

We screened three CYP2C9 substrates known to produce potentially cytotoxic reactive intermediates for CYP-dependent growth inhibition, looking for growth differences between yeast cells expressing wildtype CYP2C9 and cells expressing inactive variants. None of the substrates (benzbromarone, amiodarone, and diclofenac) have demonstrated CYP-mediated toxicity in yeast except for diclofenac, which caused demonstrated toxicity with a bacterial CYP enzyme when expressed in yeast (van Leeuwen et al. 2011). Benzbromarone is known to be

hepatotoxic, though it is unclear whether this will translate to cytotoxicity in yeast. Amiodarone was chosen for its inhibitory effects on many CYPs and ability to form a metabolite-intermediate complex with CYP3A4 (McDonald et al. 2015).

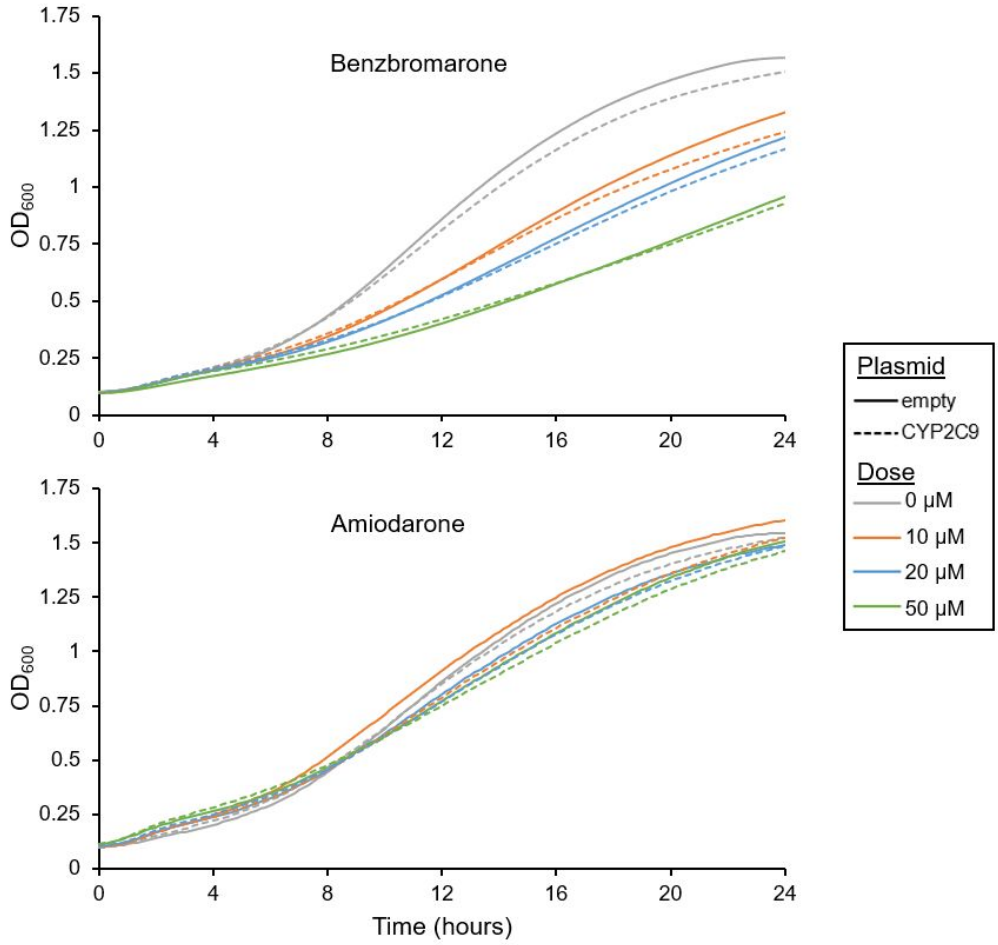

Figure 3.3. Lack of growth inhibition by benzbromarone and amiodarone in yeast expressing CYP2C9. Growth profile of YMD3289 expressing empty vector (solid lines) or CYP2C9 (dashed lines) in galactose media. Substrate concentration of benzbromarone (top) and amiodarone (bottom) indicated by color.

We expressed CYP2C9 and an empty vector control in a modified S288C derivative strain YMD3289, pre-grown in galactose media for two hours to pre-induce CYP expression, and added either benzbromarone or amiodarone. Benzbromarone showed general toxicity in high

concentrations (100 µM and above), while amiodarone did not (**Figure 3.3**). There were no significant differences in growth rate between empty vector and CYP2C9-expressing cells at any concentration of benzbromarone or amiodarone. Benzbromarone inhibition was also tested with yeast natural isolate Y12 (discussed in Chapter 2), but also did not show any CYP-dependent growth inhibition (data not shown).

We also tested CYP2C9 substrate diclofenac with WT, *2, *3, and C435H variants of CYP2C9 at five different concentrations, but found minimal differences in growth rate between variants at all concentrations (**Figure 3.4**). Noisiness in this data was likely due to precipitation issues with the substrate.

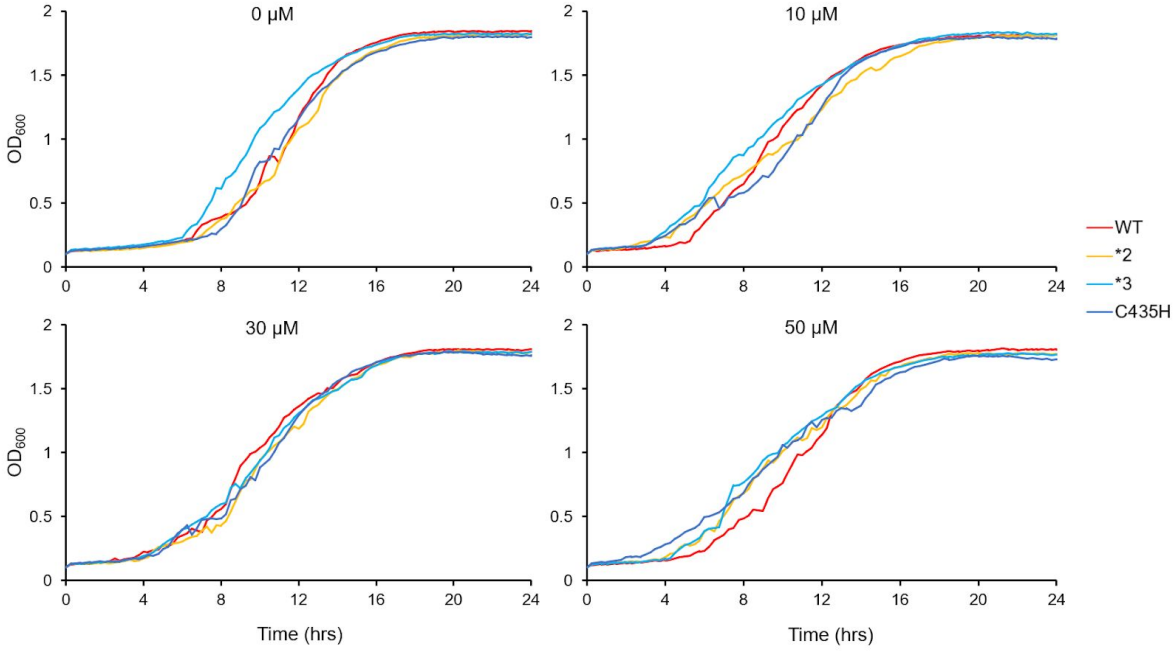

Figure 3.4. Growth profile of CYP2C9 alleles in diclofenac dosage series. Growth profile of YMD4256 expressing CYP2C9 wildtype (red), *2 (yellow), *3 (light blue), or C435H (dark blue) plasmid in galactose media. Growth in vehicle control (top left), 10 µM diclofenac (top right), 30 µM diclofenac (bottom left), or 50 µM diclofenac (bottom right).

3.2.3 Assessing CYP2C9 activity-based probes in whole yeast cells

Using ABPP in cells is possible, and labeling of CYPs with this method has even been performed in live mice (Wright and Cravatt 2007), though most common applications of ABPP are performed *in vitro*. The success of ABPP approaches depends on using the correct probe for the enzyme target. When designing CYP-specific activity-based probes, a general strategy is to modify a mechanism based inactivator by adding a terminal alkyne. This alkyne addition acts as a clickable handle that can be used to introduce a fluorophore (**Figure 3.5**). An inherent benefit of this approach is covalent labeling, which eliminates problems with product efflux. We tested existing and novel activity-based probes for CYP2C9 labeling in intact yeast cells. CYP-specific activity-based probes have been previously developed (Wright et al. 2009), and of these, aliphatic probe ABP5 (**Figure 3.7**) has the highest labeling of CYP2C9 when tested *in vitro* in supersomes.

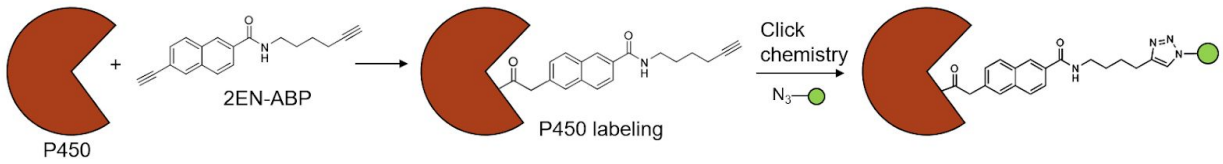

Figure 3.5. Mechanism of fluorescent P450 labeling with activity-based probe 2EN-ABP. Diagram modified from Wright and Cravatt (2007). 2-ethynylnaphthalene-derived activity-based probe (2EN-ABP) is oxidized by the P450 (red) to an electrophilic ketene that then binds to the P450 active site. Through a click chemistry step, the (unconjugated) alkyne can be covalently linked to an azide-modified fluorescent reporter (green ball).

We first tested ABP5 in our humanized yeast strain to determine whether previous *in vitro* results with this probe would replicate within yeast cells. Briefly, we induced yeast cells expressing wildtype CYP2C9 or a variant allele, incubated with NADPH and the activity-based probe, washed to remove unreacted probe, performed click chemistry to covalently bind a

fluorophore onto the CYP-linked probe, and washed to remove unlinked fluorophore. To determine whether we were seeing activity-dependent labeling, we tested an alleles series as with the fluorogenic probes, and saw much more labeling in wildtype and intermediate activity alleles compared to the null allele, C435H, thus demonstrating good dynamic range (**Figure 3.6**). However, labeling of wildtype and intermediate activity alleles was very similar with ABP5. Bimodal labeling patterns in wildtype-expressing cells was due to plasmid loss (**Figure 3.13**).

We also designed a set of novel CYP2C9 activity-based probes based on CYP2C9 mechanism-based inhibitors and (reversible) substrates (**Figure 3.7**). Tienilic acid (TA) is a well-studied mechanism-based inhibitor of CYP2C9 (López-Garcia and Mansuy 1994, Melet et al. 2003). To convert TA into an activity-based probe, tienilic acid hexynyl ester (TAHE) and tienilic acid hexynyl amide (TAHA) were synthesized. These probes were tested *in vitro* with purified CYP2C9 enzyme by performing time-dependent inhibition assays of CYP2C9-mediated diclofenac hydroxylation. With this assay, IC_{50} shift ratios greater than or equal to 1.5 generally indicate time-dependent inhibition and likely covalent linkage. ABP5, TA, TAHE, and TAHA all have IC_{50} shift ratios greater than 1.5, though the ratio is much larger for TAHE than TAHA (**Table 3.1**).

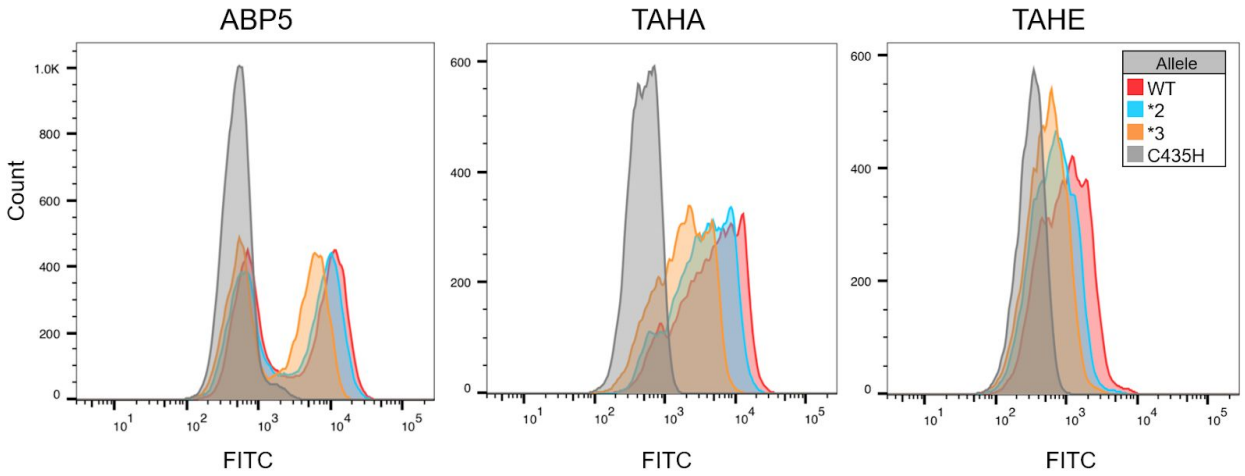

Figure 3.6. Flow cytometry of CYP2C9 alleles labeled with activity-based probes in yeast cells. Yeast strain YMD4256 expressing CYP2C9 WT, *2, *3 or C435H alleles were incubated with ABP5, TAHA, or TAHE, and labeled with a fluorophore via click chemistry. (Left) cells incubated with 20 µM ABP5 for 2 hours, (middle), cells incubated with 10 µM TAHA for 20 hours, and (right), cells incubated with 10 µM TAHE for 20 hours.

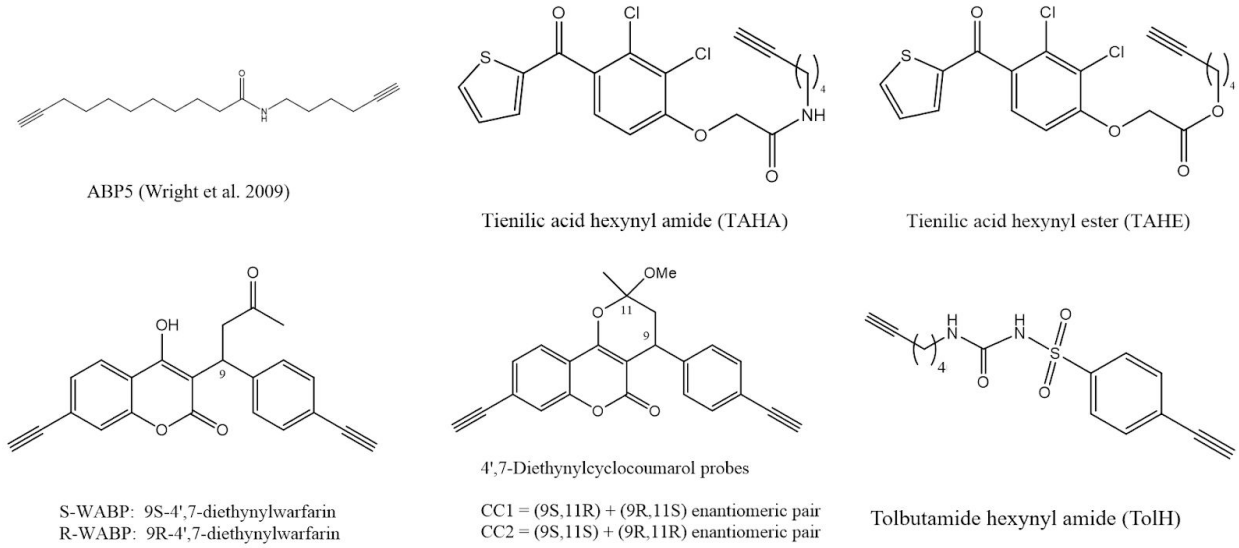

Figure 3.7. CYP2C9 activity-based probes for click chemistry. Probes tested with CYP2C9-expressing yeast cells were ABP5 (Wright et al. 2009), tienilic acid hexynyl amide (TAHA), tienilic acid hexynyl ester (TAHE), 9S-4',7-diethynylwarfarin (S-WABP), 9R-4',7-diethynylwarfarin (R-WABP), (9S,11R)+(9R,11S)-4'7-diethylnylcyclocoumarol (CC1), (9S,11S)+(9R,11R)-4'7-diethylnylcyclocoumarol (CC2), and tolbutamide hexynyl amide (TolH).

Pre-incubation	IC_{50} (μM)								
	ABP5	TA	TAHE	TAHA	S-WABP	R-WABP	CC1	CC2	TolH
(–) NADPH	4.9	7.6	3.25	1.49	0.54	2.0	4.92	2.2	37
(+) NADPH	1.4	0.49	0.28	0.82	0.64	1.7	1.56	2.2	9.2
Shift ratio	**3.6**	**16**	**11.7**	**1.82**	0.85	1.2	**3.15**	1.0	**4.0**

Table 3.1. Inhibition values of CYP2C9-mediated diclofenac hydroxylation by activity-based probes with purified enzyme. Half maximal inhibitory concentration (IC_{50}) with and without NADPH and shift ratio for seved different activity based probes and mechanism-based inhibitor tienilic acid. Shift ratios greater than 1.5 (bolded) indicate time-dependent inhibition.

These TA derivatives were tested in yeast cells with a CYP2C9 allele series as with ABP5, though we determined that a longer incubation period (20 hours) was necessary for best labeling (**Figure 3.8**). Surprisingly, TAHA showed a similar magnitude of labeling to ABP5, but TAHE had minimal labeling, even after a 20 hour incubation (**Figure 3.6**). This is somewhat surprising given the IC_{50} shift ratios of TAHA and TAHE, but this difference may be due to ester cleavage of TAHE by an endogenous factor in yeast. TAHA also showed better variant-dependent stratification of fluorescent activity than ABP5, especially with wildtype and intermediate activity alleles.

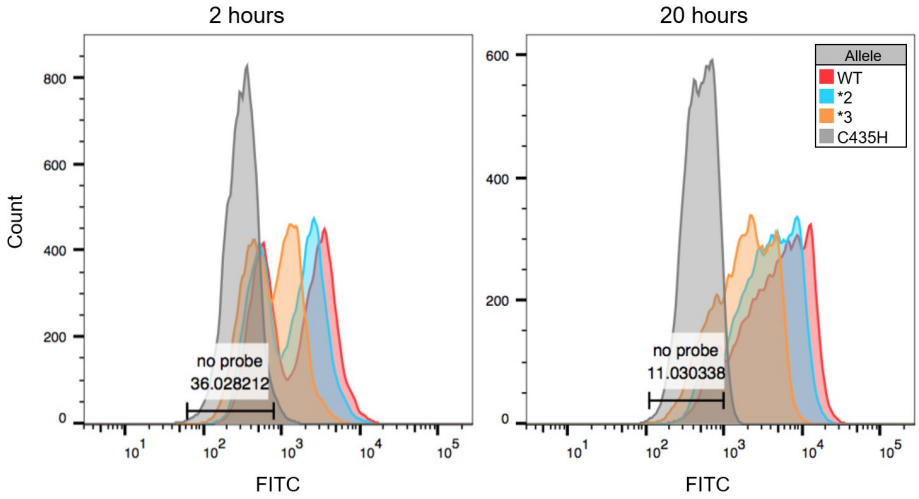

Figure 3.8. Flow cytometry of CYP2C9 alleles labeled with TAHA in yeast cells. Yeast strain YMD4256 expressing CYP2C9 wildtype (WT), *2, *3 or C435H alleles, incubated with 10 μM TAHA

for 2 hours (left) or 20 hours (right), and labeled with a fluorophore via click chemistry. No probe indicates percentage of wildtype sample overlapping with "no probe" control sample lacking TAHA.

We also tested probes derived from CYP2C9 clinical substrates. Warfarin is a widely prescribed oral anticoagulant with a narrow therapeutic range. It is administered as a racemic mixture, and S-warfarin is metabolized by CYP2C9 (Rettie et al. 1992, Flora et al. 2017). Dialkyne analogs of S- and R-warfarin were synthesized, resulting in S-WABP and R-WABP (**Figure 3.7**). However, IC_{50} shift results did not show evidence for significant time-dependent inhibition, though both S-WABP and R-WABP were strong reversible inhibitors of CYP2C9 (**Table 3.1**). Additionally, testing in yeast cells showed a minimal labeling for both probes and no differences between active and inactive CYP2C9 alleles (**Figure 3.9**), even when tested for longer probe incubations of 20 hours (not shown).

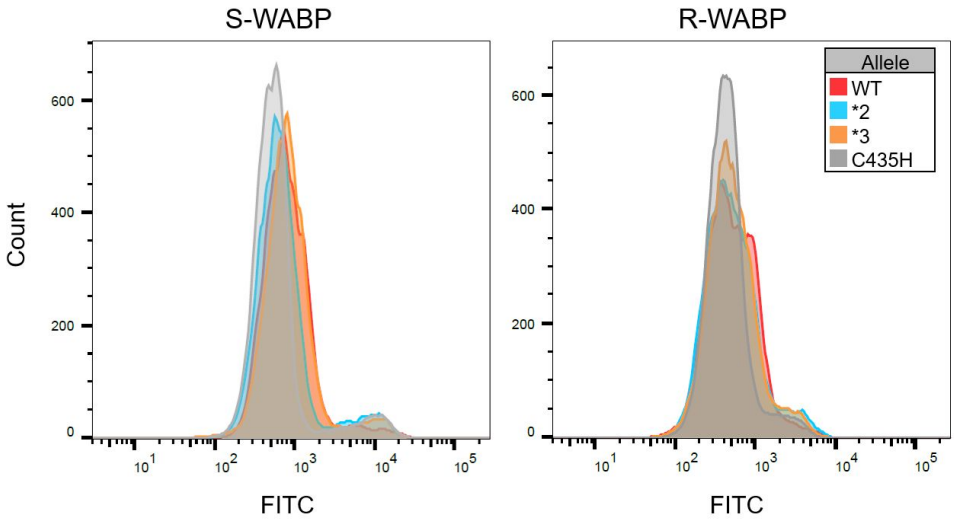

Figure 3.9. Flow cytometry of CYP2C9 alleles labeled with warfarin activity-based probes in yeast cells. Yeast strain YMD4256 expressing CYP2C9 WT, *2, *3 or C435H alleles, incubated with 20 µM S-WABP (left) or 20 µM R-WABP (right) for 2 hours, and labeled with a fluorophore via click chemistry.

Cyclocoumarols are structurally similar to warfarin, especially when warfarin is in its closed ring form, which is the conformation that binds CYP2C9 (He et al. 1999). Activity-based

probe derivatives of cyclocomarols in the form of two different enantiomeric pairs of 4'7-diethylnylcyclocoumarol, CC1 and CC2, were synthesized (**Figure 3.7**). Only CC1 showed a significant IC_{50} shift ratio indicating covalent linkage (**Table 3.1**). Results were similarly lackluster in yeast cells (**Figure 3.10**) even with both short and long incubations (data not shown). Finally, we tested a probe derived from tolbutamide, a potassium channel blocker used in the management of type 2 diabetes. Tolbutamide is used as a CYP2C9-specific substrate *in vivo* and *in vitro* (Lee et al. 2003, Kumar et al. 2006). Tolbutamide hexynyl amide (TolH) was synthesized (**Figure 3.7**) and showed an IC_{50} shift ratio of 4 (**Table 3.1**). When tested in yeast cells, the null allele showed less labeling that wildtype, but wildtype, *2, and *3 had similar labeling after four hours (**Figure 3.11**). A longer incubation with TolH did not improve labeling differences between alleles or the magnitude of wildtype labeling, which was less than labeling with TAHA or ABP5.

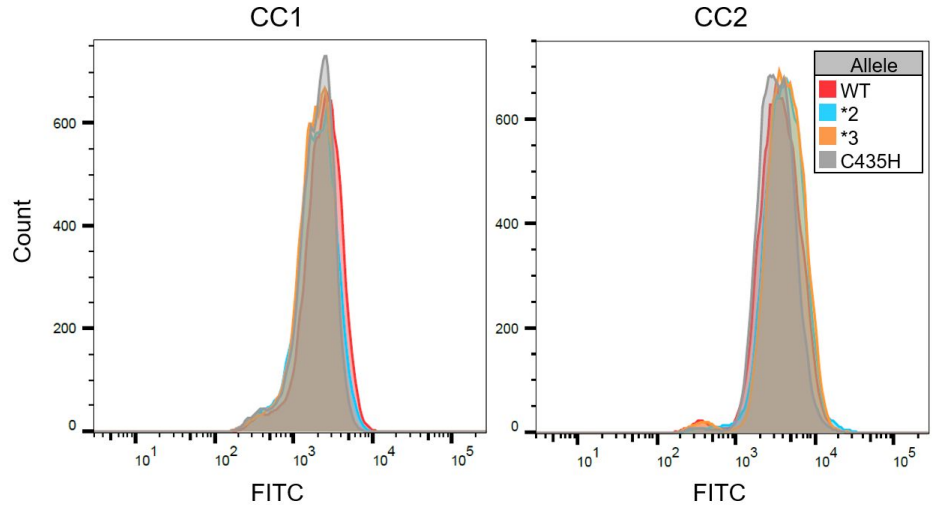

Figure 3.10. Flow cytometry of CYP2C9 alleles labeled with cyclocoumarol activity-based probes in yeast cells. Yeast strain YMD4256 expressing CYP2C9 WT, *2, *3 or C435H alleles, incubated with 20 µM CC1 (left) or 20 µM CC2 (right) for 20 hours, and labeled with a fluorophore via click chemistry.

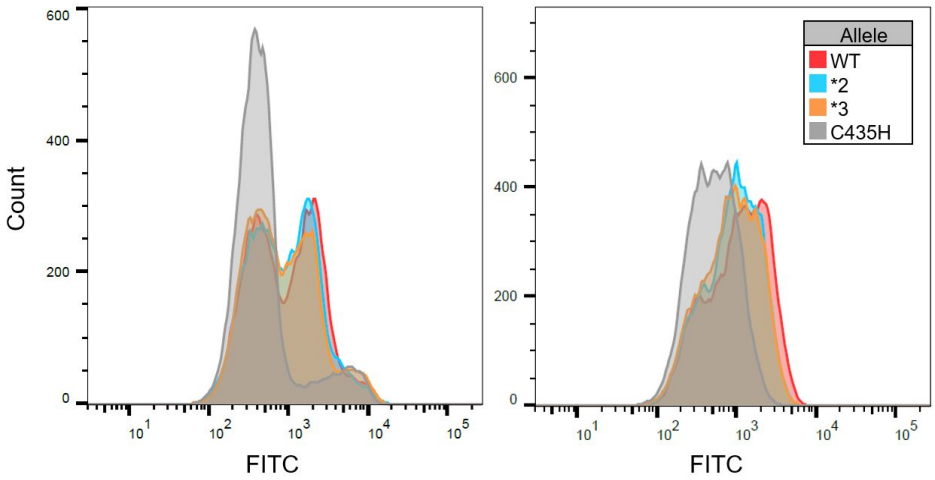

Figure 3.11. Flow cytometry of CYP2C9 alleles labeled with tolbutamide activity-based probe in yeast cells. Yeast strain YMD4256 expressing CYP2C9 WT, *2, *3 or C435H alleles, incubated with 50 µM TolH for 4 hours (left), or 10 µM TolH for 20 hours (right), and labeled with a fluorophore via click chemistry.

3.2.4 Optimizing ABPP assay in yeast cells

Developing optimal conditions for the ABPP assay in yeast cells required using a CYP2C9-specific activity-based probe, and optimizing incubation and concentration of that probe for best labeling. When labeling cells with the more general ABP5 probe, we found that the magnitude of labeling did not increase after two hours of incubation. Shorter incubation times led to more differences in labeling between wildtype and intermediate activity CYP2C9 alleles, but the overall magnitude of wildtype labeling was much less. Conversely, when labeling with the TAHA probe, much longer incubation times were necessary to obtain maximum wildtype labeling, and best results were obtained with a 20 hour incubation (**Figure 3.12A**). Varying probe dosage also improved labeling signal. In addition to TAHA, tolbutamide-derived TolH activity-based probe also showed promising results for differential labeling of wildtype and dead alleles. To try to further discriminate between intermediate activity alleles, we tested a variety of probe incubation times and concentrations. We found longer incubation times resulted

in a corresponding increase in null allele labeling, and varying probe concentration during shorter incubations did not appreciably improve the labeling differences between CYP2C9 alleles (**Figure 3.12B**). Developing a CYP2C9 probe that not only distinguishes between wildtype and null alleles, but also wildtype and intermediate activity alleles is challenging, and so far has only been successful with TAHA.

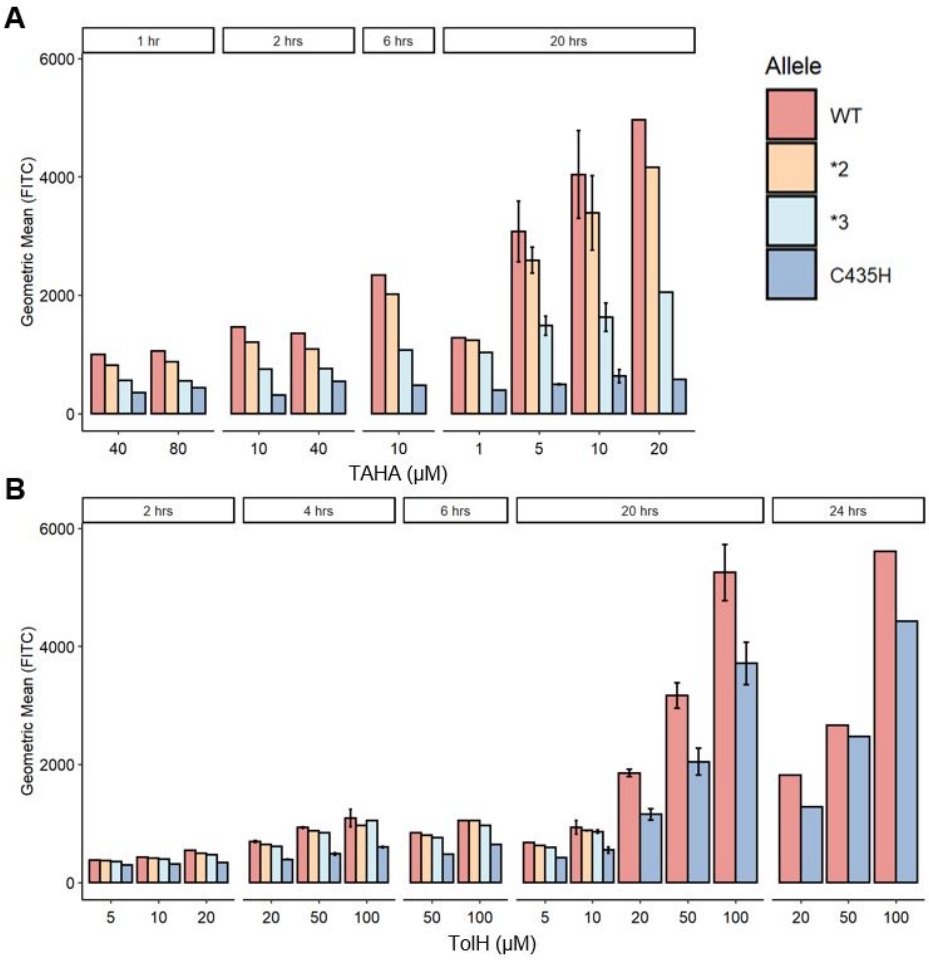

Figure 3.12. Optimization of ABPP probe incubation conditions in yeast for TAHA and TolH. Geometric mean of ABPP-labeled yeast cells (strain YMD4256) expressing CYP2C9 WT, *2, *3 or C435H alleles, labeled with TAHA (A) or TolH (B) activity-based probes. Probe concentration shown in μM, and probe incubation time in hours showed on top. Error bars represent standard deviations from biological replicates, when available.

Additionally, one other limitation of this ABPP method is toxicity to yeast cells. The type of click reaction used, copper-catalyzed azide-alkyne cycloaddition (CuAAC), causes cytotoxicity due to the Cu(I) species taken up by cells (Li and Zhang 2016). A 30 minute incubation resulted in ~0% viability with the yeast strain tested. We attempted to reduce the amount of copper used in the reaction, but using less than 1 mM $CuSO_4$ resulted in incomplete labeling and 1 mM $CuSO_4$ was toxic to yeast. Increasing the amount of THPTA ligand also did not alleviate toxicity. Ultimately, we were able to proceed with ABPP labeling despite the involved toxicity, but this is a key limitation of this method.

Finally, flow cytometry data of ABPP-labeled yeast cells highlighted the limitations of vector-based CYP2C9 expression. Even though yeast cells are grown under drug selection to maintain plasmid, we observed cell populations with ~50% plasmid loss after galactose induction (**Figure 3.13**, left panel). After a two hour incubation with ABP5 and subsequent click labeling, two clear populations of cells with and without plasmid (and thus with and without CYP2C9 protein to label) are evident with vector-based expression. Conversely, when CYP2C9 alleles are stably integrated into the genome, only one population is present (**Figure 3.13**, right panel). Sanger sequencing of these sorted cells populations confirmed that only the wildtype-labeled population contained wildtype plasmid.

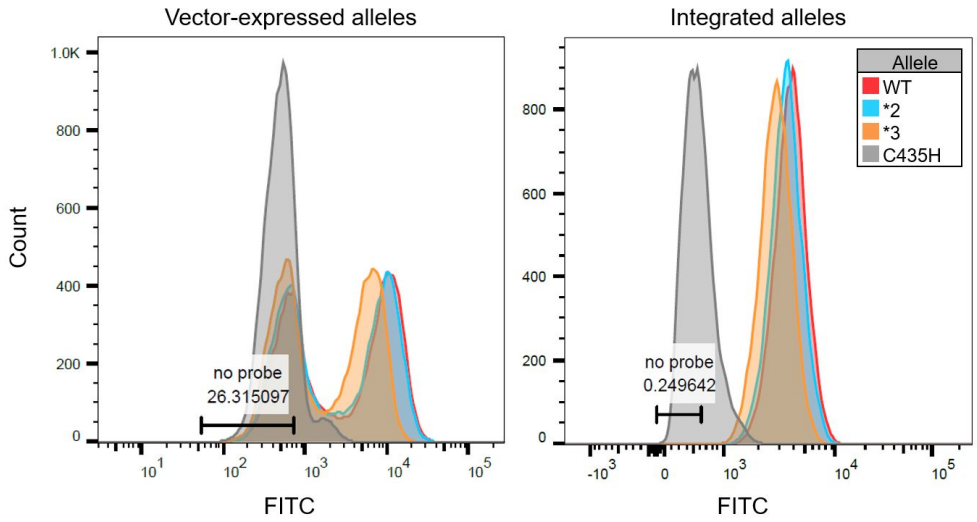

Figure 3.13. Flow cytometry of CYP2C9 alleles expressed from plasmids or integrated into the genome. Left) Yeast strain YMD4255 expressing CYP2C9 WT, *2, *3 or C435H alleles from a low copy plasmid, incubated with 20 µM ABP5 for 2 hours and labeled with a fluorophore via click chemistry. Right) Yeast strain YMD4255 with genome-integrated CYP2C9 WT, *2, *3 or C435H alleles, incubated with 20 µM ABP5 for 2 hours and labeled with a fluorophore via click chemistry. No probe indicates percentage of WT sample overlapping with "no probe" control sample lacking ABP5.

3.3 Discussion

To develop a pooled activity assay for CYP2C9 in yeast, we took a variety of approaches including testing fluorogenic substrates in pooled assays, assessing different substrates for growth inhibition assays, and determining the best CYP2C9 activity-based probes for activity-based protein profiling (ABPP) in whole cells. Of these, the third approach was the most successful because it is a covalent labeling approach, which avoids the issue of metabolite diffusion between cells in a pooled sample.

We tested a fluorogenic substrate, BOMCC, in a mixed population of wildtype and null CYP2C9 alleles and found evidence of metabolite diffusion between cells and into the surrounding media. Ultimately, an approach using a fluorogenic substrate can only be performed when cells expressing different alleles are physically separated. One potential way of doing this

is to use a droplet-based microfluidics approach, which is naturally suited to high-throughput screening approaches. Droplet-based technologies typically rely on a fluorescent readout of enzyme activity, and can be coupled with fluorescence-activated sorting for screening libraries of variants (Theberge et al. 2010). Droplet technologies typically rely on fluorescent readouts of activity or other function, but are not immune from issues of substrate diffusion. In particular, fluorogenic substrates derived from coumarin (such as BOMCC) are quickly transported between compartments in droplet-based microfluidics systems, which reduces their effectiveness (Woronoff et al. 2011). Thus, developing a CYP2C9-specific droplet-based microfluidics assay may rely on screening many fluorogenic substrates or developing new ones. Microfluidics-based droplet methods have been successful in screening libraries in yeast (Huang et al. 2015, Huang et al. 2018), but developing a droplet-based assay would likely require extensive substrate testing and optimization.

One method of measuring CYP activity in a pooled fashion is to use a growth inhibition assay. Given the right substrate, CYP-mediated products will cause cytotoxicity in cells and can be an indirect readout of activity. This requires finding a substrate that leads to CYP-dependent toxicity (through a known or unknown mechanism), but also one where the cytotoxic products are not released into the surrounding media and cells. Focusing on CYP2C9, we tested benzbromarone, amiodarone, and diclofenac for CYP-dependent growth inhibition in a lab strain of *S. cerevisiae*. Benzbromarone showed general cytotoxicity in *S. cerevisiae*, but did not show evidence of CYP-dependent toxicity. Amiodarone, conversely, did not show much toxicity in our strain, even though it has been shown to have antifungal properties in many fungal species due to disruption of calcium homeostasis (Courchesne et al. 2002). Finally, diclofenac also showed no

evidence of CYP-dependent toxicity, even though this substrate had been previously shown to have a metabolism-mediated toxicity, albeit with a bacterial CYP enzyme (van Leeuwen et al. 2011). This negative result may be due to the relatively low concentration of CYP2C9 per yeast cell and/or the residence time of diclofenac within the cell before efflux. In the future, we would like to screen more substrates such as tienilic acid and losartan to see if we can uncover one that is amenable to a CYP-dependent growth inhibition assay in *S. cerevisiae* for CYP2C9 or other CYP enzymes.

We developed an ABPP approach to profile CYP activity in whole yeast cells in a pooled format. Key to this method was using a new activity-based probe that is a modified form of tienilic acid, a known covalent inhibitor of CYP2C9 (Jean et al. 1996). With this probe, yeast expressing human CYP2C9 were labeled in an activity-dependent manner, even distinguishing between wildtype and intermediate activity variants. Since labeling is covalent, this assay can be performed in a pooled format without the leakage or diffusion of metabolites. When combined with FACS and high-throughput sequencing, it can be used to determine the activity of each variant within a CYP library pool of thousands. Though we found that TAHA produced the best results with CYP2C9, we also observed activity-dependent labeling in yeast cells with ABP5 (Wright et al. 2009), and to a lesser extent, a tolbutamide activity-based probe, TolH. Activity based probes derived from S- and R-warfarin and related cyclocoumarol substrates showed little evidence of time-dependent inhibition of CYP2C9. In order to develop a clickable activity-based probe, a CYP substrate must first be modified to have a free alkyne (or azide) for conjugation with a fluor-labeled azide (or alkyne) for CuAAC (click chemistry). This may reduce the affinity of CYP for the substrate, necessitating additional optimization.

One main caveat of CuAAC reactions (and thus ABPP) is that it is generally toxic to cells due to high copper levels. In our system, yeast cells were non-viable after a 30 minute incubation with the copper-containing click cocktail, and shorter incubations resulted in incomplete labeling. Though we determined that this was an acceptable limitation in our system, modifying the ABPP approach to retain viability would be a beneficial improvement. In mammalian cells, viability has been successfully retained during CuAAC and relies on ligand-accelerated CuAAC reactions (Hong et al. 2011), but in our system may require other approaches such as strain-promoted Cu(I)-free azide-alkyne cycloaddition (SPAAC) or chelation-assisted CuAAC (Li and Zhang 2016). One other limitation of this system, plasmid loss, complicates the interpretation of flow cytometry data, but is ultimately not an issue with multiplexed assays since cells that have lost the variant plasmid will not be sequenced due to the design of amplification primers.

Overall, we developed a cell-based activity assay that is capable of labeling a mixed population of CYP2C9 variants. Though we optimized this assay for CYP2C9 labeling, this approach is easily amenable to other CYP isoforms. Wright et al. (2009) developed a suite of clickable probes, and several of these probes (ABP5 and ABP6) label CYP2D6 well. Thus we expect success with other CYPs, though expression level differences may have a large effect on the effectiveness of labeling certain CYPs (for example, CYP2C19, which is expressed at a much lower level than CYP2D6 in yeast cells).

3.4 Materials and Methods

Growth media and culturing techniques: Growth media and culturing techniques for *E. coli* and *S. cerevisiae* as described in Chapter 2 Materials and Methods. Genotype information for strains used in this chapter listed in **Table 2.2.**

Site directed mutagenesis:. CYP2C9 variants were generated using inverse PCR (Jain and Varadarajan 2014). Using oligonucleotide pairs whereby the forward primer contains a NNK at the 5' end of the sequence, point mutations were generated by amplifying vector pRS41KGAL-CYP2C9-HA with KAPA HiFi DNA Polymerase (KAPA Biosytems KK2601). Variant fragments were treated with T4 polynucleotide kinase (NEB M0201) and ligated with T4 DNA ligase (NEB M0202) before transforming chemically competent *E. coli* cells with the ligated products and miniprepping (Qiagen). Variants were verified with Sanger sequencing. Primers sequences are as follows: CYP2C9 R144C (*2): CJA059F: tgtGTGCAAGAAGAAGCTCGTTGTCTTGTGGAA, CJA059R: ATCCTCAATTGATCTTTTACCCATACCGAAATT; CYP2C9 I359L (*3): CJA054F: ttgGACTTGTTGCCTACCAGCTTGCCGCATGCA, CJA054R: GTAACGTTGAACTTCATGTACGACTGCATCAGT, CYP2C9 C435H: CJA073F: catGTAGGCGAAGCCCTGGCGGGAATG, CJA073R: TATTCTTTTTCCCGCTGAGAATGGCATAAAGT.

Strain engineering: A portion of the pRS41KGAL-CYP2C9-HA (or variant allele) vector containing both *GAL1-hCYP2C9-HA* and *KanMX* was amplified via PCR with HIS3 flanking homology, digested with DpnI, and used to transform the strain YMD4255 by selecting for growth on media supplemented with 200 μg/ml G418.

Fluorogenic substrate assay in whole cells: To measure fluorescence turnover in a microplate reader, yeast cells carrying a CYP vector were induced as described above. After harvesting, cells were rinsed twice with 0.5 mL 50 mM KPi buffer, pH 8, and 1 OD of cells was resuspended in buffer and mixed together with 50 μM 7-Benzyloxymethyloxy-3-cyanocoumarin (BOMCC) OR 7-ethoxymethyloxy-3-cyanocoumarin (EOMCC) and 200 μM NADPH in 150 μL final incubation volume. Each sample was done in parallel with a no substrate control. Sample fluorescence (excitation: 410nm, emission: 460nm, gain: 70% unless otherwise specified) was recorded every 5 minutes on a BioTek Synergy H1 microplate reader at 37°C for 200 min with shaking.

To measure fluorescence efflux, after assay completion, cells were pooled from each sample, spun down, supernatant was saved, and cells were washed with 0.5 mL 50 mM KPi buffer, pH 8. Wash step was repeated twice, after which cells were resuspended in 0.5 mL 50 mM KPi buffer, pH 8. Fluorescence of supernatant and cells was measured on a BioTek Synergy H1 microplate reader (excitation: 410nm, emission: 460nm, 70% gain).

To measure fluorogenic substrate turnover using flow cytometry, cells were induced as described above, and 2 OD of cells was spun down and washed with 0.5 mL 50 mM KPi buffer, pH 8. Cells were resuspended in KPi buffer, and 200 μM NADPH and 50 μM EOMCC was added to each sample. Samples incubated overnight at 30 °C, then washed five times 0.5 mL KPi buffer, and resuspended in KPi buffer. Cell fluorescence measured on a LSRII on the BV 450 channel, 20,000 events each.

IC50 shift assays in human liver microsomes: Inhibition of CYP2C9-mediated diclofenac 4'-hydroxylation by ABP5 and novel CYP2C9 activity-based probes in human liver microsomes was performed as described in McDonald et al. (2015).

Activity-based protein profiling in yeast cells using flow cytometry: Yeast cultures were grown as described above to induce CYP expression, and for each sample, 1 OD of overnight yeast culture was collected via centrifugation at 4000rpm for 2 min, washed with 0.5 mL of PBS by resuspension and centrifugation, and resuspended in 100 µL PBS:0.1% saponin (w/v). Each sample was pre-incubated with 2mM NADPH (Sigma N1630) at 37°C for 20 mins. All samples except a 'No probe' control were treated with activity-based probe at desired concentration and incubated with rotation at 37°C for 20hrs (unless otherwise specified) to form activity-dependent CYP2C9-probe adducts. Samples were collected via centrifugation at 4000rpm for 2 min and washed three times with 0.5 mL PBS. Samples were resuspended in 100µL PBS:0.1% saponin (w/v) and incubated at room temperature for 20 mins. Cells were washed twice via resuspension in 1mL PBS and centrifugation at 10,000g for 5 minutes. Cells were resuspended in 100µL copper-catalyzed azide-alkyne cycloaddition (CuAAC) reaction buffer to append a fluorophore reporter (2x concentrations: 10 µM CF488A picolyl azide (Biotium #92187), 2 mM $CuSO_4$ (Sigma C8027), 4 mM THPTA (Sigma 762342), 6 mM ascorbic acid (Sigma A7631) in PBS) and vortexed vigorously to mix. Samples were incubated in the dark at room temperature for 30 minutes and collected by centrifugation as above. Cells were washed five times by resuspension/centrifugation in 0.5 mL PBS, resuspended in 1 mL PBS, and stored at 4°C up to 1 day. Labeled cells were analyzed using a BD LSRII and forward and side scatter, and FITC (488 nm excitation; 530/30 nm detection filter) parameters were

collected for 20,000 events. Flow cytometry data was collected using FACSDiva (BD Biosciences) and analyzed using FlowJo (Ashland, OR). Fluorescence was calculated from the FITC geometric mean of gated single cells.

Chapter 4: Massively parallel functional profiling of *CYP2C9* variants

---

The work described here was performed by Clara Amorosi and collaborators. Library construction was conducted by Elisa Wong and Kate Sitko. Kate Sitko and Melissa Chiasson performed the VAMP-seq experiments. Probe synthesis and LC-MS assays were performed by Matt McDonald. Maitreya Dunham, Doug Fowler and Allan Rettie contributed to the writing and editing of this chapter. This chapter as written is under revision and will be submitted as a manuscript under the same name.

The vast amount of genetic variation being discovered far outpaces the capacity for interpreting variant effects. For example, the lack of variant functional annotation in genes that dictate drug response prevents most variants from being used to inform drug choice and dosing. We addressed this problem in a particularly important pharmacogene: *CYP2C9*. *CYP2C9* encodes a cytochrome P450 enzyme responsible for metabolizing drugs, including warfarin, phenytoin, and flurbiprofen, and variants in this gene affect the efficacy of these and other drugs. A variety of multiplexed assays of variant effect have been developed to quantify the effect of genetic variation on gene expression, protein function and protein activity. However, few multiplexed assays are capable of directly measuring enzyme activity in cells. Therefore, we developed a yeast-based activity assay to test thousands of variants in a pooled fashion, called Click-seq. We measured activity scores for 6,142 single missense variants of CYP2C9 and found that the regions most important to activity are in the core of the protein and involved in heme

binding. Moreover, 65% of missense variants have significantly decreased activity suggesting altered drug metabolism *in vivo*, and we provide activity scores for 289 missense variants already documented in the human population but with no previous functional annotation. Additionally, we performed a second deep mutational scan of CYP2C9 in a human cell line using variant abundance by massively parallel sequencing (VAMP-seq). Abundance scores were obtained for 6,370 single missense variants, 4,421 of which also have an activity score. Consequently, we could determine whether variant loss of function was due to decreased abundance, activity, or both. Our CYP2C9 functional scores correlate with gold standard *in vitro* metabolic assays when tested on a subset of individual variants. Click-seq can be extended to other CYP enzymes, and will lead to advances in adverse drug response prevention by providing clinical guidance for patients carrying both currently known and yet-to-be discovered alleles of *CYP2C9*.

4.1 Introduction

Recent sequencing efforts have resulted in an avalanche of new variants, many of which are variants of uncertain significance (VUSs), variants identified through genetic testing whose functional significance is unknown. VUSs hamper the implementation of precision medicine as they must be classified into functional categories before they can be used to inform clinical decisions. Over half of the missense variants in ClinVar (Landrum et al. 2014) are VUS (Starita et al. 2017). In population databases such as GnomAD (Karczewski et al. 2020), the majority of variants have unknown functional effects. This is a particular problem in the field of pharmacogenomics, where only functionally annotated variants can be used to guide dosing decisions and predict adverse drug reactions.

Clinical implementation of pharmacogenomic knowledge has lagged behind the discovery of genetic interactions with many drugs. In the US, adverse drug reactions (ADRs) are a leading cause of hospitalization and death (Lazarou et al. 1998), and 30% of these are predicted to be caused by inter-individual variability in drug metabolizing enzymes and other drug related genes (Budnitz et al. 2006). Understanding genetic variability in drug-metabolizing enzymes is crucial to predicting ADRs, which can occasionally have fatal consequences (Koren et al. 2006) and cost U.S. hospitals up to 4 billion dollars annually (Lazarou et al. 1998). Additionally, being able to predict how a patient's genotype will affect the metabolism of a given drug can increase drug efficacy. Genetics predicts drug response for a subset of important drugs, and implementing genotype-guided dosing with known alleles is a proven strategy for improving patient outcomes (Goulding et al. 2015).

One of the most important PGx enzyme families are the cytochromes P450 (CYPs), which are responsible for metabolism of the majority of drugs. Included in this family is CYP2C9, which metabolizes the widely prescribed oral anticoagulant warfarin, as well as a number of other drugs including tolbutamide and phenytoin. CYP2C9 functional status determines 15-20% of the variation in warfarin dose (Duconge et al. 2009), and genotype-guided warfarin dosing can improve patient treatment in some situations (Pirmohamed et al. 2013). *CYP2C9* has a handful of common alleles (MAF > 0.5%) of known activity, and also has a large number of individually rare (MAF < 0.5%), yet collectively common alleles (Gordon et al. 2014). Human CYP alleles are named according to the star (*) system (Sim and Ingelman-Sundberg 2010) and curated by the PharmVar consortium (Gaedigk et al. 2018). *CYP2C9* currently has 61 star alleles, 55 of which are single amino acid mutations

(www.pharmvar.org). Knowing the functional consequence of these and yet-to-be discovered alleles will help improve dosing of drugs cleared by CYP2C9. Thus, there is a need for a large-scale experimental effort to comprehensively characterize CYP2C9 variants.

Efforts to functionally characterize CYP variants at scale have been hindered by the low-throughput nature of classic biochemical assays and the inconsistencies of computational variant effect predictors. Therefore, we adopted a deep mutational scanning (DMS) approach to measure the functional capabilities of thousands of CYP2C9 variants. DMS applies a functional selection to a library of protein variants, stratifying variants with high and low function or activity (Fowler and Fields 2014, Chiasson et al. 2019). DMS approaches have been successful with BRCA1 (Findlay et al. 2018, Starita et al. 2015) and many other genes (Sun et al. 2020, Matreyek et al. 2018). Variations of DMS may measure a specific property such as protein abundance; this is the case with Variant Abundance by Massively Parallel Sequencing (VAMPseq, Matreyek et al. 2018) which uses a fusion EGFP reporter coupled with flow sorting. To date, no one has performed a DMS on the activity of a human CYP enzyme, even though there is a clear need for such comprehensive functional annotation.

We developed a multiplexed, yeast-based activity assay (Click-seq) and used it to measure the activity of thousands of CYP2C9 variants. Additionally, we leveraged VAMP-seq (Matreyek 2018) to measure CYP2C9 abundance in a human cell system. Measuring both activity and abundance revealed the extent to which each variant's function is impacted by reduced activity and/or reduced stability. From these two assays, we provide activity and/or abundance scores for a combined total of 8,091 (or 87%) of all possible single missense variants of CYP2C9, the overwhelming majority of which did not previously have functional annotation.

This dataset will be of great utility to clinicians as a key source of evidence when presented with VUS.

4.2 Results

We characterized a library of CYP2C9 single amino acid variants using Click-seq, leveraging activity-based protein profiling and fluorescence-activated cell sorting (FACS). We additionally generated a second CYP2C9 library to test CYP2C9 abundance in mammalian cells, using an existing approach, VAMP-seq. With two sets of functional scores, we show which regions of CYP2C9 are crucial for protein abundance and which are solely important for CYP2C9 activity.

4.2.1 Using activity-based protein profiling as a readout of CYP activity in yeast

We developed a massively parallel assay of CYP activity using activity-based protein profiling (ABPP) (Cravatt et al. 2008). This technique, called Click-seq, uses CYP-selective, covalent probes to modify the enzyme, click chemistry to label probes with a fluorophore, FACS to separate cells according to their labeling, and high-throughput sequencing of sorted cells to score each variant (**Figure 4.1a**). This method directly measures activity by quantifying activity-dependent inhibition of CYP enzyme after a period of incubation. CYP-specific ABPP probes have been developed previously (Wright and Cravatt 2007, Wright et al. 2009), but prior work has focused on *in vitro* assays (commonly CYP-rich microsomal preparations), rather than cell-based methods. We modified existing ABPP assays to work with intact yeast cells in a pooled format. We synthesized a new activity-based probe, TAHA, that is a modified form of tienilic acid, a known covalent inhibitor of CYP2C9 (Jean et al. 1996). This probe showed better

labeling than general P450 probe ABP5 from Wright et al. (2009). Additionally, to improve recombinant CYP activity, human P450 accessory proteins were integrated into a modified laboratory strain (see Methods) resulting in a humanized strain.

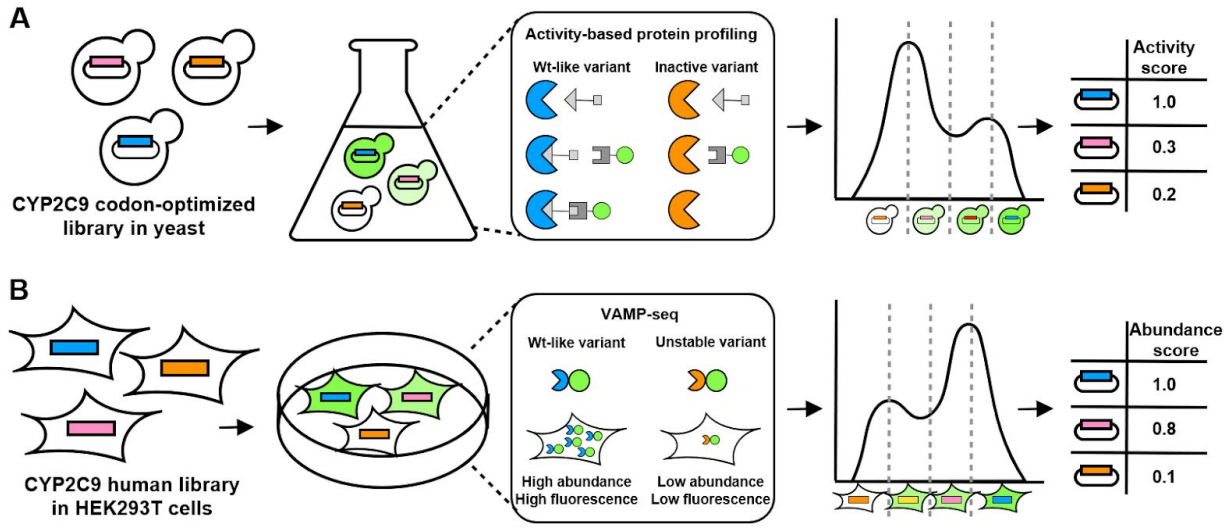

Figure 4.1. Overview of CYP2C9 library selection schemes. In A), a humanized yeast strain is transformed with a library of codon optimized CYP2C9 variants, labeled using activity-based protein profiling, resulting in a range of fluorescence levels, and sorted into four bins using fluorescence-activated cell sorting. In B), using a method called VAMP-seq (Matreyek et al. 2018), HEK 293T cells are transfected with a CYP2C9 library where each variant is expressed with an EGFP fusion, resulting in a range of fluorescence according to variant stability. This library is flow sorted as in A). For both A) and B), sequencing is used to determine variant distribution across bins, resulting in activity and abundance scores, respectively.

In order to demonstrate that Click-seq accurately reflects enzyme activity, we cloned individual CYP2C9 variants of known activity and compared labeling levels to wild-type CYP2C9. We found that, as expected, CYP2C9 *2 (R144C) and *3 (I359L) have decreasing levels of fluorescent labeling (**Figure 4.2**), and a catalytically inactive variant, C435H (a negative control where the heme-coordinating cysteine is mutated), has labeling comparable to background levels. After validating with alleles of known activity, we constructed a barcoded,

site-saturation mutagenesis library of CYP2C9, from positions 2 to 490, codon optimized for yeast expression. This library covers 6,542 of the 9,780 possible single amino acid variants (67%), with 105,372 barcodes. The CYP2C9 activity library was labeled using tienilic acid-derived probe TAHA, flow sorted into bins and DNA collected from each bin was amplified, sequenced, and analyzed to determine relative variant activity. Four replicate sorts from four separate outgrowths were performed on this library.

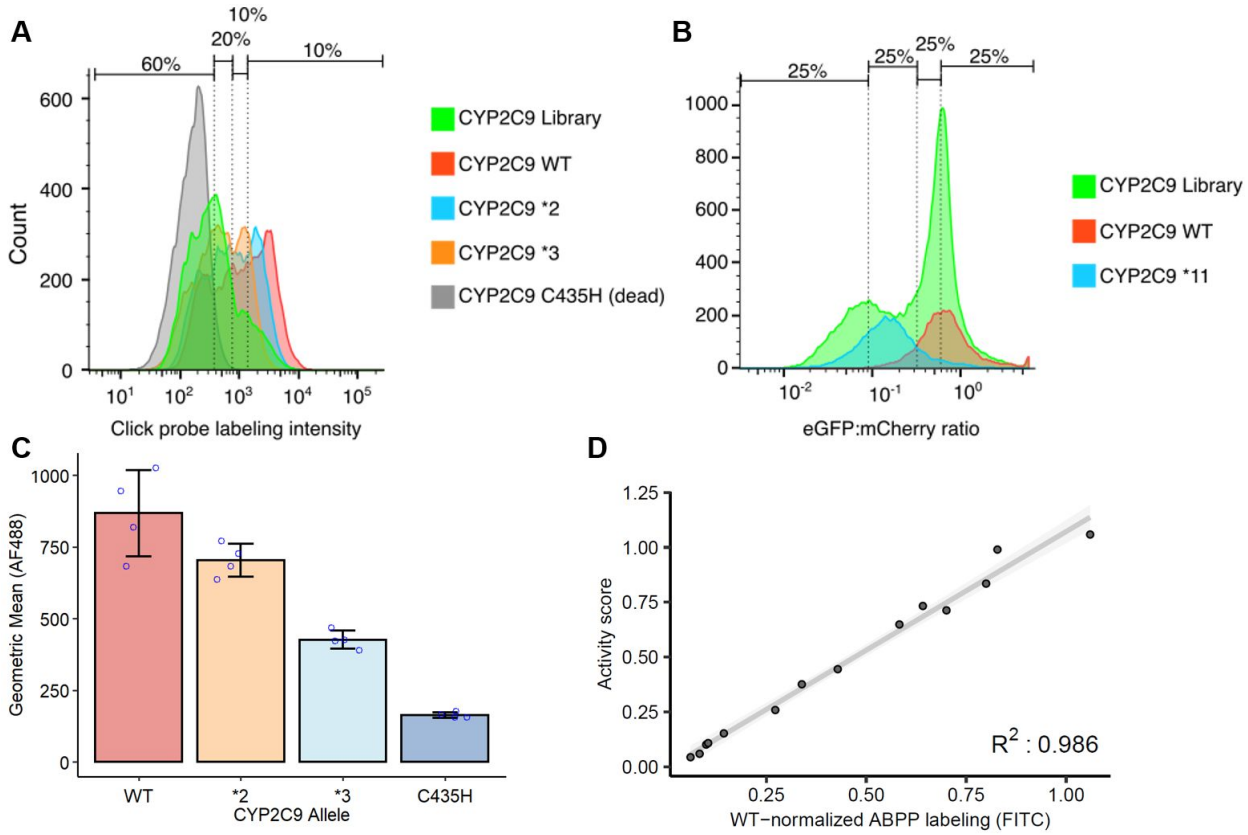

Figure 4.2. CYP2C9 activity and abundance library sorts and validation. A) Flow cytometry of yeast-expressed and ABPP labeled CYP2C9 WT (red), reduced activity alleles (*2 and *3, blue and orange), null allele (C435H, grey) and CYP2C9 variant library (green). Smoothed histograms shown, each sample represents ~20,000 cells. B) Flow cytometry of CYP2C9 WT (red), destabilizing allele (*11, blue), and CYP2C9 EGFP fusion library expressed in HEK293T cells (green). Smoothed histograms of EGFP:mCherry ratios shown. Approximate library bins shown as percentages at top. C) Geometric mean of ABPP-labeled CYP2C9 alleles, individual replicates shown as blue points, and error bars show standard deviation. D) WT-normalized ABPP labeling (FITC normalized fluorescence) for 14 CYP2C9

variants, expressed in the humanized yeast strain and labeled separately. Individual variants were labeled using the same ABPP protocol as the pool assay. Scatter plot and linear regression of activity score (pool score) versus individual variant ABPP labeling.

After filtering the data from the yeast activity library sequencing, we calculated activity scores (see Methods) for 6,524 single variants, of which 6,142 were missense, 131 were nonsense, and 250 were synonymous (**Figure 4.3**). Activity scores are normalized to median nonsense and synonymous variant scores such that a score of 0 represents nonsense-like activity and a score of 1 represents wildtype-like activity. Variant activity scores correlate very well between the four replicate sorts (mean Pearson's r=0.92, see Appendix **Figure A.3**). Strikingly, 47% of missense variants have an activity score less than 0.5, and when we binned activity scores into activity classes, we found that 65% of missense variants have significantly decreased activity (see Methods).

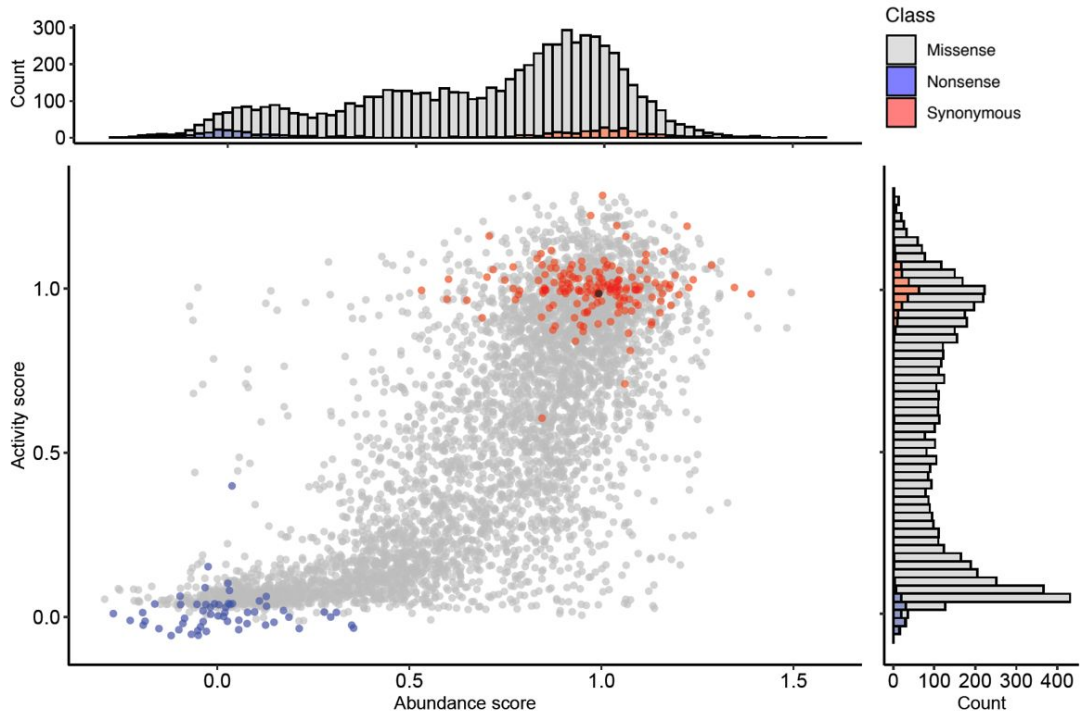

Figure 4.3. Correlation of CYP2C9 activity and abundance scores. Scatter plot of CYP2C9 activity and abundance scores. Shows a total of 57 nonsense variants (blue), 165 synonymous variants (red), and

4,421 missense variants (grey). Wildtype shown in black. On top, stacked histogram of abundance score colored by type of variant. On right, stacked histogram of activity score colored by type of variant.

To internally validate our assay, we generated and tested 14 individual CYP2C9 variants and compared them to pool-generated activity and abundance scores. For the activity assay, individual variant labeling measured via flow cytometry matched well with activity scores (**Figure 4.2D**) after bin weight optimization (see Methods). To show that our large-scale activity scores determined with a non-clinical substrate are representative of CYP2C9 activity towards conventional drugs, we performed gold-standard LC-MS assays of S-warfarin 7-hydroxylation and phenytoin 4-hydroxylation activities with 14 CYP2C9 variants that span the full range of activity scores, using microsomes prepared from yeast-expressed alleles (**Figure 4.4**). Activity scores are correlated well with individual variant S-warfarin turnover (Pearson's r=0.874, Spearman's ρ=0.895). The same is true with phenytoin turnover (Pearson's r=0.764, Spearman's ρ=0.87). Both of these CYP2C9 drug substrates have highly similar activity levels across the 14 variants tested (Pearson's r=0.965, Spearman's ρ=0.979). Interestingly, the G442S variant, which has an activity score of 1.06, shows increased turnover of both S-warfarin (128% of wildtype turnover rate) and phenytoin (176% of wildtype turnover rate). Additionally, we tested these variant microsomes with a fluorogenic substrate, BOMCC, and found similar correlation with activity scores (Pearson's r=0.801, Spearman's ρ=0.829).

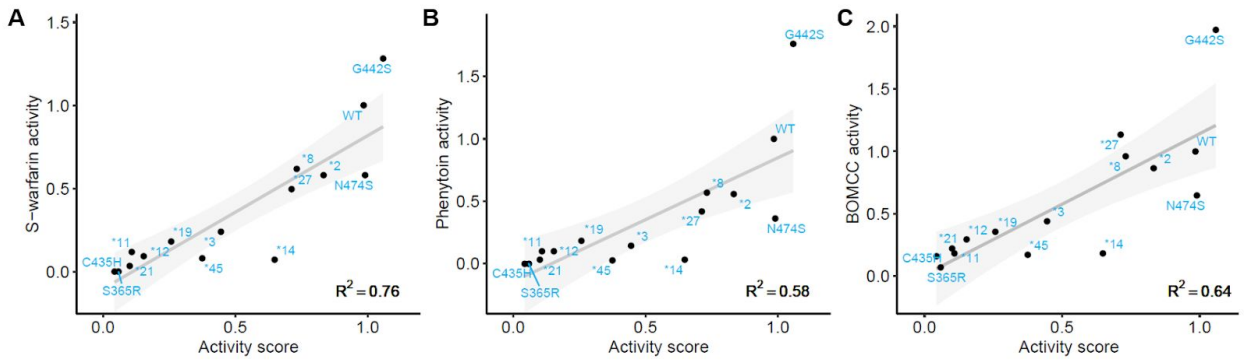

Figure 4.4. Comparison of CYP2C9 activity scores with gold-standard activity assays on yeast microsomes. Scatterplots of CYP2C9 activity scores plotted against individually tested CYP2C9 alleles. In A-C) individual alleles were expressed in the humanized yeast strain used in the pooled assay, and yeast microsomes were harvested from these individual strains. In A), LC-MS was used to determine the rate of S-warfarin 7-hydroxylation. In B), LC-MS was used to determine the rate of phenytoin 4-hydroxylation. In C), the conversion of BOMCC to CHC (fluorescent) was monitored using a plate reader. All activities are shown normalized to wildtype rates.

4.2.2 CYP2C9 pooled abundance assay in HEK 293T cells.

We recently developed a method, VAMP-seq (Matreyek et al. 2018), that enables measurement of steady-state protein abundance in cultured human cell lines using fluorescent reporters. We applied VAMP-seq to CYP2C9, fusing eGFP C-terminally, and from the same construct expressing mCherry via an internal ribosomal entry site (IRES) to control for cell-to-cell differences in expression. We then conducted a pilot experiment, in which we measured the ratio of eGFP to mCherry of both WT CYP2C9 and CYP2C9 R335W (*11), a known destabilized variant (**Figure 4.2B**). Because CYP2C9 R335W is destabilized, the protein misfolds, and protein quality control degrades both CYP2C9 and the fused eGFP, leading to lower eGFP intensity (**Figure 4.1B**). After validating that R335W indeed had lower eGFP signal, we constructed a barcoded, site-saturation mutagenesis library of CYP2C9, from positions 2 to

490. This library covers 8,310 of the 9,780 possible single amino acid variants (85%), with 78,740 barcodes.

We expressed this library in HEK 293T cells using a lentiviral serine integrase landing pad (Matreyek et al. 2020, Matreyek et al. 2018). After selection, recombinant cells were assayed based on eGFP:mCherry ratio and sorted into four quartile bins (**Figure 4.2B**). Bins were deeply sequenced, and abundance scores were calculated based on relative variant frequency across bins and normalized to synonymous and nonsense mutations as was done with activity scores. Three replicate sorts were performed on this library. Variant abundance scores showed distinct, separable distributions of synonymous and nonsense variants, with missense variants spanning the range of scores (**Figure 4.3**). After filtering, we assigned variant scores and classifications to 6,821 single variants, of which 6,370 were missense, 189 were nonsense, and 261 were synonymous, and the three replicates correlated well (Pearson's $r = 0.789$, Spearman's $\rho = 0.75$). Compared to the activity scores, only 29% of missense variants have an abundance score less than 0.5, while only 37% of missense variants had a significantly decreased abundance class.

4.2.3 Mechanism of CYP2C9 variant loss of function.

We generated both activity and abundance scores (**Figure 4.5**) for 4,644 CYP2C9 variants (4,421 missense), and observed a strong correlation between the two sets of scores (**Figure 4.3**, Pearson's r=0.76, Spearman's ρ=0.755). We observe an abundance threshold at a score of ~0.5 below which variants have a very low activity (median activity score of 0.098 for these variants), suggesting that variants with abundance below this level, differences in ABPP assay signal are too small to detect. Conversely, variants with abundance scores greater than 0.5 have a wider range of activity scores. The overall positive trend between abundance and activity

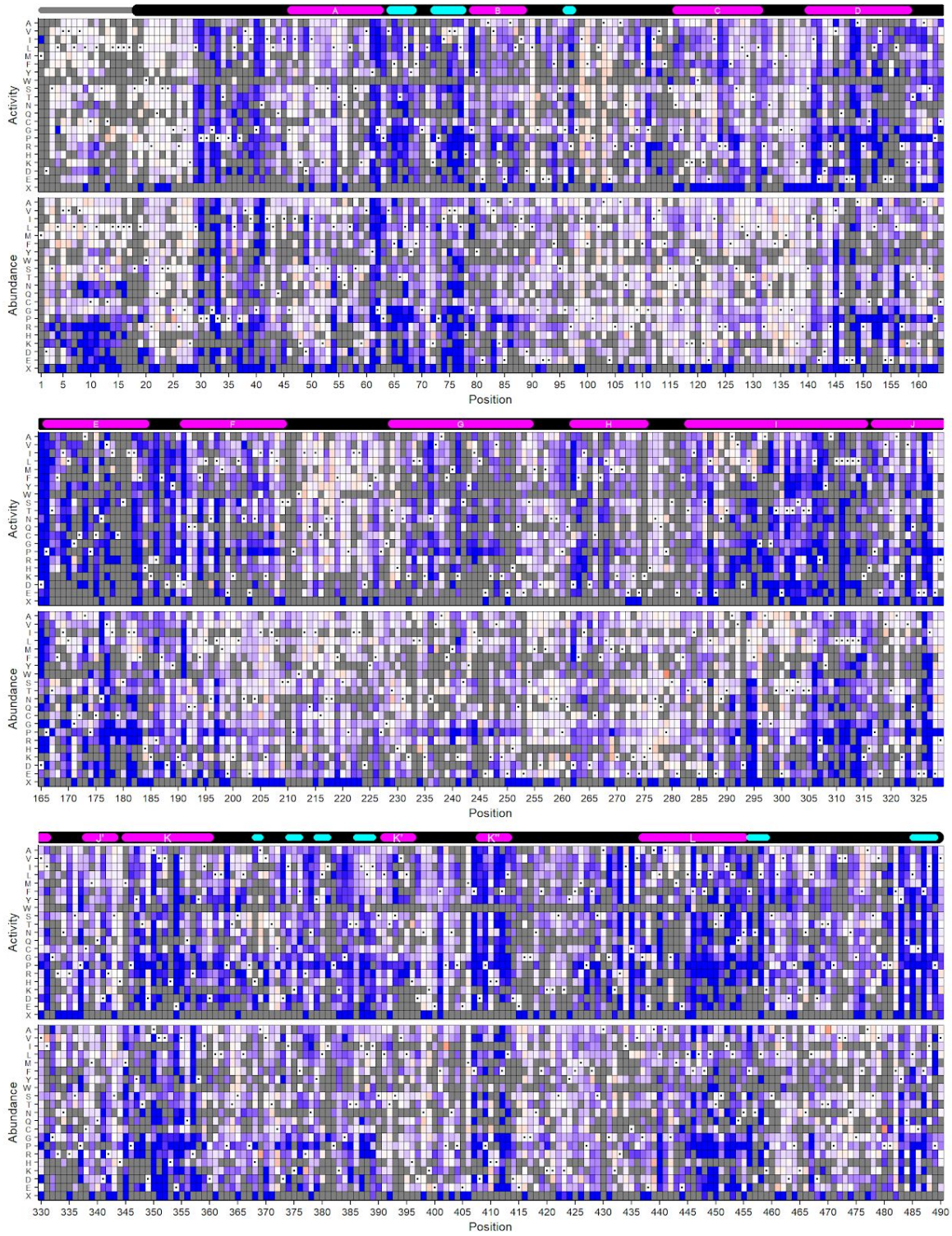

Figure 4.5. CYP2C9 activity and abundance score heatmaps. On top, secondary structure of CYP2C9, with alpha helices in magenta and beta sheets in cyan. Helices are labeled by letter name. Below,

heatmaps of CYP2C9 activity and abundance scores. WT amino acids denoted with a dot, and missing data is shown in grey. Scores range from nonfunctional (blue) to wt-like (white) to increased (red).

scores tells us that 1) using engineered yeast as a heterologous CYP expression system largely recapitulates protein behavior in human cells, and 2) a large portion of variants have low activity because they are less abundant. Since there is no normalization to protein expression per cell, the yeast activity scores reported are each a combination of both variant activity and variant stability.

With these two sets of scores, we can identify variants that abolish activity but not abundance. As an example of this, I359L (*3) has an activity score of 0.45 but an abundance score of 0.78, meaning that this variant reduces activity but is still fairly abundant. L361I (*55) is a more extreme example of this, with an activity score of 0.57 and an abundance score of 0.95. Using a basic metric (whether a variant has an activity score and an abundance score less than 0.5), we see that protein stability leads to reduced total activity for at least 26% of missense variants with both scores, and this is likely a conservative estimate.

4.2.4 Structural insights from CYP2C9 scores

Microsomal CYP enzymes are composed of an N-terminal ER-transmembrane domain and a large, cytoplasmic catalytic domain. P450 enzymes are diverse at the sequence level but share a common structure including 12 major helices, labeled A through L, and four beta sheets, labeled β1 through β4 (Gay et al. 2010). CYP2C9 has been crystallized with warfarin (Williams et al. 2003) and flurbiprofen (Wester et al. 2004) as well as with other substrates (Maekawa et al. 2017) and unliganded. We first mapped median positional activity and abundance scores onto the CYP2C9 structure (**Figure 4.6**). We focused in particular on positions with the lowest abundance scores and found these positions cluster into two distinct regions: positions in and directly abutting β sheet-1 and core-facing positions in helices D, E, I, J, K, and L. Both of these regions

are highly conserved across CYPs and are composed of buried, hydrophobic residues (Hasemann et al. 1995), substitution of which leads to destabilization and degradation. In addition, substitutions in β sheet 1 may disrupt distal side chains that coordinate with the central heme iron (Arendse and Blackburn 2018).

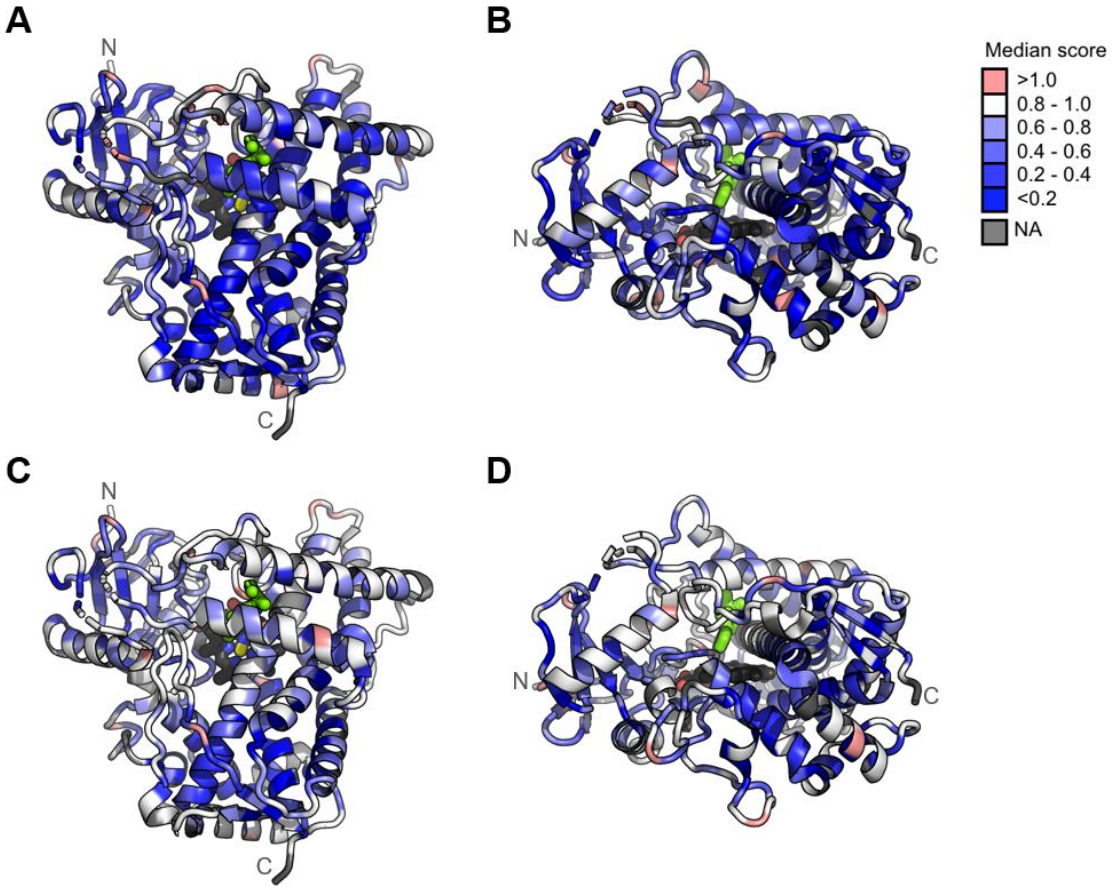

Figure 4.6. Structural features important to CYP2C9 function. CYP2C9 structure (PDB: 1r9o) colored by median score at each position. In A) and B), activity scores shown, and in C) and D) abundance scores shown. Median scores are binned as depicted in legend, and missing positions shown in grey. Heme colored by element (carbon:black, nitrogen:blue, oxygen:red, iron:yellow), and substrate (flurbiprofen) colored bright green. A rotated viewpoint is shown in B) and D).

We expect that the highly conserved regions important for abundance discussed above will also be intolerant to mutation in our activity assay, but that the active site and binding pocket

of CYP2C9 will only show intolerance to mutation in the activity data. To further explore this, we performed hierarchical clustering on these scores by variant position. Looking into the six main position clusters (**Figure 4.7**), we see that mutations in cluster 3 fall into the universally non-tolerated category and that these positions are grouped in the residues of the β sheet-1 region and in the inner core-facing positions of helices D, E, I, J, K. Slightly more tolerated are mutations in cluster 2 and 1, which are located in inner residues, though cluster 2 is much more tolerated in the abundance assay than the activity assay. In particular, mutations in the I helix, which runs along the heme and contains a central bulge for O_2 binding, are highly deleterious to activity. Conversely, mutations in cluster 4 are almost universally tolerated and these positions are located on the surface of the protein. Mutations in cluster 6 behave similarly to cluster 4 but are slightly less tolerated. And finally, mutations in cluster 5 are highly enriched in the N terminus, including the transmembrane domain which is not shown in the crystal structure. Mutations in the transmembrane domain (residues 1-20) have little effect on activity (with an average median score of 0.949), but have a larger effect on abundance, especially charged and polar mutations (**Figure 4.5**). This suggests that CYP2C9 does not require correct ER localization when expressed in yeast, but we have not yet confirmed this with microscopy.

Finally, there are several important highly conserved motifs that are recapitulated in our datasets. First and most important is the heme binding motif FXXGXXXCXG at positions 428 - 437 (Danielson 2002), which contains the heme-coordinating cysteine at position 435. Cys435 is extremely intolerant to mutation in our activity assay (all 18 missense mutations at this site have “nonsense-like” or “possibly nonsense-like” scores) but shows less effect on abundance (only 4 of the 9 missense mutations at this site even have a “decreased” abundance score). This is also

true for Gly431, Gly437, and somewhat true for Phe428. In addition to the known heme binding motif preferences, we see that a basic residue is preferred at position 433, and also slightly preferred at position 432. Another CYP motif is the proline-rich PPGP motif at positions 30-33, which is in the linker (or hinge) region. Mutations in this motif result in large decreases in both activity and stability, and the average of median activity scores for the PPGP motif is 0.18, which is quite low. This region's intolerance to mutation has been shown previously for other CYP2C isoforms (Kemper 2004), but we see from our data that Leu29 of CYP2C9 also is extremely important to activity, with a median score of 0.165. Additionally, Arg108 has been shown to be critical for binding negatively charged substrates (Dickmann et al. 2004) and for binding flurbiprofen in particular (Wester et al. 2004). However, the probe used in our activity assay is an amide and not acidic, so variants at Arg108 may not impact activity in our assay. Indeed, the median activity and abundance score at this position are 0.943 and 0.910, respectively. Finally, Arg97 is known to be important to CYP2C9 enzyme stability and heme propionate binding (Dickmann et al. 2004) and mutations at this site are not tolerated in the activity assay, but are slightly better tolerated in the abundance assay.

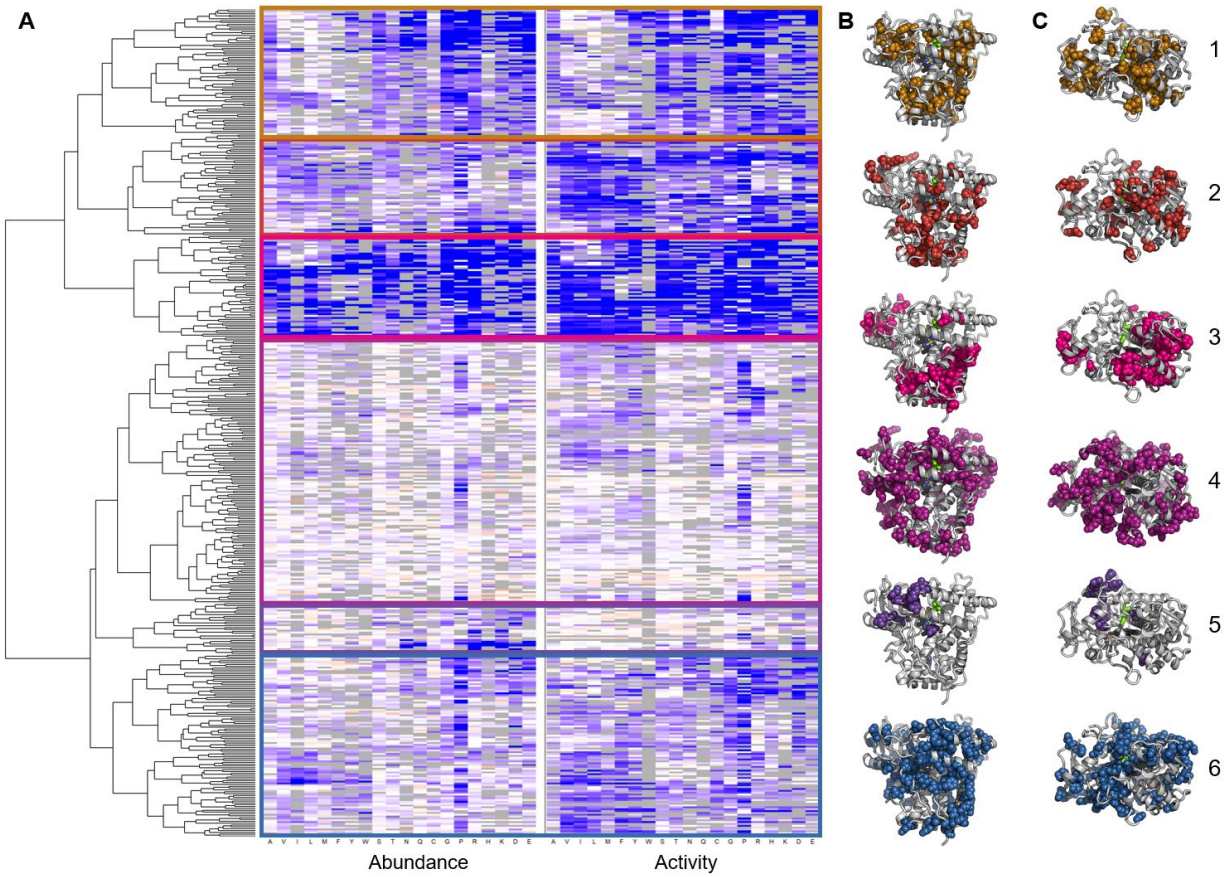

Figure 4.7. Clustering of activity and abundance score by position. In A), dendrogram and heatmaps of CYP2C9 activity and abundance score clustered by position. Heatmaps colored as in Figure 4.5. Colored boxes on heatmaps indicate the six major clusters (numbered), and correspond to the colors shown on the CYP2C9 structures in B) and C). In B) and C), the positions that correspond to each of the six clusters are shown as spheres in the corresponding color on the CYP2C9 crystal structure (PDB: 1r9o). Two different views are shown in B) and C).

4.2.5 Human CYP2C9 variants with predicted clinical impact.

The Clinical Pharmacogenetics Implementation Consortium (CPIC) reviews *in vitro* and *in vivo* evidence and provides clinical functional recommendations for CYP2C9 and other pharmacogenes (Relling and Klein 2011). CPIC allele biochemical functions (either "normal function", "decreased function", or "no function") are provided for 32 of the 55 CYP2C9 single amino acid star alleles (Theken et al. 2020). These CPIC allele function classes are generally

concordant with our CYP2C9 activity classes (**Figure 4.8**), and the few cases where our activity classes do not match CPIC classes are generally due to alleles with limited or inadequate functional evidence, as determined by CPIC (see Appendix Table A.3). We additionally curated CYP2C9 variants from the population database GnomAD (Karczewski et al. 2020), which does not have functional annotations. We find 466 missense variants, 229 of which are singletons. We classify 340 of these variants based on activity and find that 58.8% (200 variants) have "decreased" or "possibly decreased" activity, and 9.7% (33 variants) have "nonsense-like" or "possibly nonsense-like" activity. Synonymous and stop-gained variants are also present in GnomAD data, and these are enriched in the "wt-like" and "nonsense-like" activity classes, respectively (Appendix Figure A.6). The large number of CYP2C9 missense variants in GnomAD with significantly decreased activity indicates the clinical utility of this dataset.

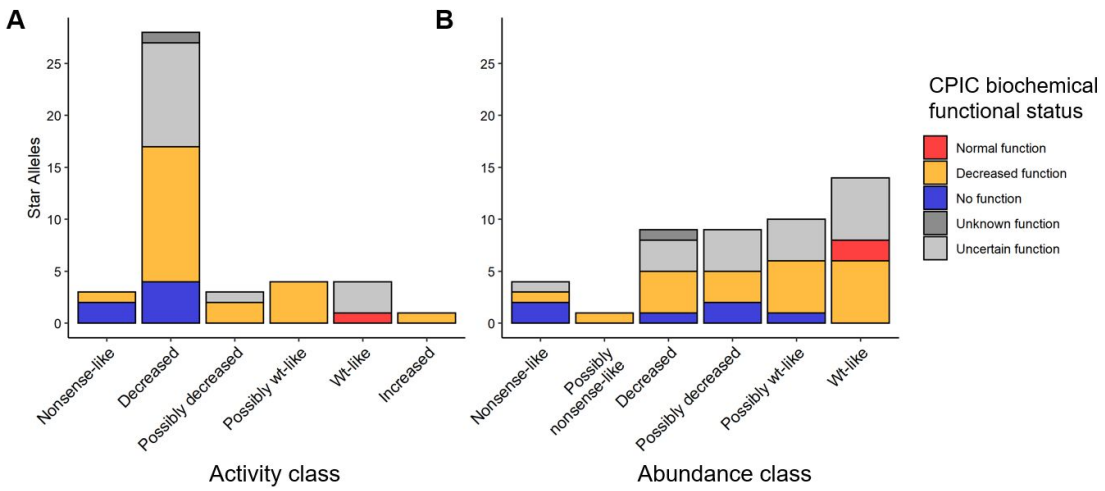

Figure 4.8. CPIC star allele function status across score classes. Stacked bar plot of A) activity and B) abundance classes colored by CPIC biochemical functional class status. CPIC classes are taken from NSAID clinical functional status recommendations (Theken et al. 2020).

Conversely, only 240 variants are classified as having increased function with a maximum activity score of 1.3, though the clinical usefulness of a (homozygous) variant with a

potential 30% increase in activity is unclear, especially since there are no documented CYP2C9 increased activity alleles. However, 20 of these 240 increased activity variants are present in GnomAD, and one variant, I434F (*59), is present in the PharmVar database, warranting further investigation.

4.3 Discussion

Microsomal cytochromes P450 are well-studied metabolic enzymes, and many small-scale functional characterizations of CYP2C9 have been performed (Niinuma et al. 2014, Wang et al. 2014, Dai et al. 2013), with the largest of these comprising of 100+ CYP2C9 variants and profiling CYP2C9 abundance (Zhang et al. 2020). Overall, previous studies of CYP2C9 variants have tested only a small fraction of the possible single mutations, and do not test yet-to-be-discovered CYP2C9 alleles. Therefore, we developed Click-seq, a high-throughput yeast activity assay, performed two deep mutational scans of CYP2C9, and generated activity and abundance scores for a combined total of 8,091 missense variants, or 87% of the possible missense variants in CYP2C9.

Our functional scores reflect known features of microsomal CYPs, including the highly conserved heme-binding motif, and helices close to the heme and substrate binding pocket. We find that CYP2C9 stability explains a large portion of enzyme function, as decreased activity variants often also have decreased abundance scores. In general, the regions most important to CYP2C9 abundance cluster in the hydrophobic core of the globular domain. Residues involved in heme coordination and binding are less important to protein abundance than activity, indicating that heme insertion, a process which is not fully understood, is not necessarily stabilizing (Correia et al. 2011). However, protein abundance does not fully capture enzyme

activity. For example, while Zhang et al. classify 34 CYP2C9 variants as tolerated based on abundance, we show that 10 of these have decreased activity. Thus protein abundance cannot provide the full picture of CYP2C9 function.

Our results are highly repeatable across biological replicates and validate well when tested against individual variants. Additionally, our activity scores are concordant with CYP2C9 star allele functional status recommendations put forth by CPIC (Theken et al. 2020). In addition to CYP2C9 alleles of known function, we have generated functional scores for hundreds of CYP2C9 alleles present in population databases that currently lack functional annotation. There are 466 unique CYP2C9 single missense variants in GnomAD, and we generated activity scores for 340 of these.

We hope these datasets will be a resource for improving genotype-based dosing and also for improving our understanding of CYP biology, but these assays do have some limitations. First, in both systems we express CYP2C9 as a cDNA using inducible promoters, so we cannot measure splicing defects or transcriptional regulation. CYP2C9 is constitutively expressed in the liver, but is also inducible via a number of substrates (Miners and Birkett 1998, Zanger and Schwab 2013). We also cannot discern the impact of protein interactions such as with CYP accessory proteins cytochrome P450 reductase (CPR) and cytochrome b5, or with other CYP enzymes. Additionally, due to our flow cytometry binning strategy, we suspect that our yeast assay is saturated at increased activity levels, so we would likely need to re-sort our library with a modified binning strategy to detect variants with significantly increased activity (this can be seen with the G442S variant, which has a “wt-like” activity score but shows 130% and 180% wildtype activity in individual tests). Finally, Click-seq measures CYP2C9 activity using a single

substrate, so we cannot say how many variants may exhibit substrate-dependent effects, for which there is some evidence (Lee et al. 2002). However, our activity scores correlate well with S-warfarin and phenytoin activity, indicating that our functional scores should be informative for a larger set of substrates. In the future we would like to re-test our library with a range of activity-based probes to identify variants that result in substrate-dependent changes in function.

The data presented here can be used to inform drug dosing. As preemptive genotyping efforts increase in clinical settings, we will continue to find new variants, and some of these will already have functional annotations due to this study and other MAVE studies. In the future, the yeast-based activity assay developed in this paper can be leveraged to examine other CYPs important to human drug metabolism such as CYP2D6 and CYP2C19, both of which have also been successfully expressed in yeast previously and for which ABPP probes have been designed (Wright et al. 2009). Expanding the repertoire of CYP deep mutational scans will allow us to investigate differences between CYP isoforms that are key to human drug metabolism and pharmacogenomics.

4.4 Materials and Methods

General reagents. Unless otherwise noted, all chemicals were obtained from Sigma Aldrich Chemical Co. (St. Louis, MO) and all enzymes were obtained from New England Biolabs. Tienilic acid, 6-hydroxywarfarin-d_5, 7-hydroxywarfarin-d_5 and 4-hydroxyphenytoin-d_5 were synthesized according to published protocols (Rademacher et al. 2012). Hex-5-yn-1-amine was purchased from GFS Chemicals (Powell, OH).

Growth media and culturing techniques. *E. coli* were cultured at 37°C in Luria broth. Yeast culture media was prepared according to the following recipes. YP: 1% yeast extract, 2%

peptone. YPD: 1% yeast extract, 2% peptone, 2% (w/v) glucose. Yeast were cultured at 30 °C and unless otherwise specified, all yeast transformations were performed using the protocol described in Gietz and Schiestl 2007.

Yeast cells carrying a *CYP2C9* or variant plasmid were induced as follows: a single colony was inoculated into 5mL YPD media supplemented with 200 μg/ml G418 and grown overnight with rotation. This culture was diluted 1:50 into fresh YP media containing 2% (w/v) raffinose and supplemented with 200 μg/ml G418 and grown for at least two cell doublings. Cultures were then inoculated to OD 0.0125 into fresh YP media containing 2% (w/v) galactose and 200 μg/ml G418 and collected after 7 doublings the following day.

All cell culture reagents were purchased from ThermoFisher Scientific unless otherwise noted. HEK 293T cells (ATCC CRL-3216) and derivatives thereof were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 100 U ml^{-1} penicillin, and 0.1 mg ml^{-1} streptomycin. Cells were induced with 2.5 ug mL^{-1} doxycycline. Cells were passaged by detachment with trypsin-EDTA 0.25%, and cells were prepared for sorting by detachment with versene. All cell lines tested negative for mycoplasma.

Yeast strain engineering. Details on construction of humanized yeast strain and other strain engineering detailed in section 2.4 Materials and Methods.

CYP2C9 yeast codon-optimized variant library construction in *S. cerevisiae.* Missense variants were generated using an inverse PCR-based site-directed saturation mutagenesis approach (Jain and Varadarajan 2014). Using oligonucleotide pairs whereby the forward primer contains a NNK at the 5' end of the sequence (Integrated DNA Technologies), point mutations were generated using a KAPA HiFi DNA Polymerase (KAPA Biosytems

KK2601) and a *CYP2C9* template sequence. To generate the template vector, the *S. cerevisiae* codon-optimized *CYP2C9* sequence from pRS41K*GAL1pr::hCYP2C9-HA* was cloned into pHSG298 (Güldener et al. 1996) using restriction sites SalI and XbaI. After performing inverse PCR for each amino acid position in *CYP2C9*, the variant constructs were verified by gel electrophoresis prior to quantification by Qubit fluorometry (Life Technologies) and pooling at equimolar ratios. Variant fragments were treated with T4 polynucleotide kinase (NEB M0201) and ligated with T4 DNA ligase (NEB M0202) before transforming electrocompetent *E. coli* cells (NEB C2989K) with the ligated products and midiprepping (Qiagen). Next, the library was subcloned back into the low-copy pRS41K*GAL1* vector using restriction sites SpeI and SalI and ligated products were used to transform electrocompetent *E. coli* cells (NEB C2989K) and midiprepped (Qiagen). The variant library was digested with SalI and barcoded using 18-bp random sequences introduced via Gibson assembly downstream of the variant sequences as in Ahler et al. 2019. The barcoded library was used to transform electrocompetent *E. coli* cells (NEB C2989K) and midiprepped (Qiagen). The size of the barcoded library was estimated using colony counts to be 280,000. To reduce library size, the barcoded library was again used to transform electrocompetent *E. coli* cells (NEB C2989K), bottlenecked, and midiprepped (Qiagen). The size of the barcoded library was estimated using colony counts to be 42,000.

To determine more accurate library barcode counts, 2 PCR replicates each using 1.5 µg of plasmid extracted library were amplified using custom barseq primers CJA120/CJA138 using KAPA2G Robust HotStart ReadyMix (Sigma 2GRHSRMKB) with the following conditions: 95°C for 3m, 5 cycles of 95°C for 15s, 60°C for 15s, 72°C for 15s, and 72°C for 1m, then purified using AMPure XP beads (Beckman Coulter A63880) at 1:1 ratio (beads:DNA). The

purified products were amplified using primers CJA135 and JS486 or JS487 using KAPA2G Robust HotStart ReadyMix with the following PCR conditions: 95°C for 3m, 10 cycles of 95°C for 15s, 65°C for 15s, 72°C for 15s, and 72°C for 1m, and then gel extracted using the QIAquick Gel Extraction Kit (Qiagen) and quantified by Qubit fluorometry (Life Technologies). PCR replicates were pooled at equimolar ratios and deep sequenced on an Illumina NextSeq500 to determine the number of barcodes present. Briefly, forward and reverse reads were merged with Pear (Zhang et al. 2014), barcodes were counted with Enrich2 (Rubin et al. 2017), and barcodes with less than 10 reads were removed, resulting in a total of ~160,000 unique barcodes in the *CYP2C9* library, for an average of 17x coverage.

The barcoded *CYP2C9* library was used to transform the humanized yeast strain YMD4256, using the standard high-efficiency LiAc procedure mentioned above. Four independent transformations were pooled to generate a library stock of OD_{600} 5.7, equivalent to an average of 11x coverage (independent transformants) for each of the 160,000 independent barcoded variants. The latter estimate is based on the assumption that each yeast cell harbors one CYP2C9 variant and that all growth rates are similar. Library stocks were stored at -80°C in 25% (v/v) glycerol.

CYP2C9 human library construction in HEK293T cells. A gBLOCK with an optimized sequence for human CYP2C9 was ordered from IDT. It was then cloned into the vector pHSG298 (Clontech). Saturation mutagenesis primers were designed for each codon in CYP2C9 from positions 2 to 490 and ordered resuspended from IDT. Forward and reverse primers for each position were mixed at 2.5 mM, and used in a PCR reaction with 125 pg of

pHSG298-CYP2C9, 5% DMSO, and 5 µL of KAPA Hifi Hotstart 2X ReadyMix. PCR products were visualized on a 0.7% agarose gel to confirm amplification of the correct product.

PCR products were then quantified using the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen) using DNA control curves done in triplicate. To pool, a total amount of DNA for each reaction was calculated that maximized the volume to be drawn from the lowest concentration PCR product. Pooled PCR products were cleaned and concentrated using Zymogen Clean and Concentrate kit and then gel extracted. The pooled library was phosphorylated with T4 PNK (NEB), incubated at 37°C for 30 minutes, 65°C for 20 minutes, and then 4° indefinitely. 8.5 µL of this phosphorylated product was combined with 1 µL of 10X T4 ligase buffer (NEB) and 0.5 µL of T4 DNA ligase (NEB) to make a 10 µL overnight ligation reaction. This reaction was incubated at 16°C overnight.

The overnight ligation was then cleaned and concentrated (Zymogen) and eluted in 6 µL of ddH2O. 1 µL of this ligation was then transformed into high efficiency E. coli (NEB C3020K) using electroporation (settings: 2 kV). Each reaction contained 1 µL of ligation (or ligation control or pUC19 10 pg./uL) and 25 µL of E. coli. 975 µL of pre-warmed SOC media was added to each cuvette after electroporation, transferred to a culture tube, and recovered at 37°C, shaking for 1 hour. At 1 hour, 1 and 10 µL samples from all cultures were taken and plated on appropriate media (LB + kanamycin for ligation and ligation control; LB + ampicillin for pUC19), the remaining 989 µL was used to inoculate a 50 mL culture (+ kanamycin). Plates and 50 mL culture were incubated at 37°C overnight (shaking for 50 mL culture). Colonies on plates were then counted, and counts were used to calculate how many unique molecules were transformed to gauge coverage of the library. 50 mL culture was spun down and midiprepped.

To transfer the library from pHSG298 to the recombination vector, the pHSG298 library and recombination vector were digested with MluI and SphI for 1 hour at 65°C. The library and cut vector were then gel extracted. The library was then ligated with the cut vector at 5:1 using NEB T4 ligase, overnight at 16°C. The ligation was heat inactivated the next morning, clean and concentrated with the Zymo kit. Another high efficiency transformation was performed the same as described above, except this ligation was plated on LB + ampicillin (antibiotic switching strategy). Plates and 50 mL culture were incubated at 37°C overnight (shaking for 50 mL culture). Colonies on plates were then counted, and counts were used to calculate how many unique molecules were transformed to gauge coverage of the library. 50 mL culture was spun down and midiprepped.

To barcode individual variants, plasmid library harvested from midiprep was digested with AgeI-HF at 37°C for 1 hour, 65°C for 20 minutes. Barcode oligos were ordered from IDT, resuspended at 100 uM, and then annealed by combining 1 µL each of primer with 4 µL CutSmart Buffer and 34 µL ddH2O and running at 98°C for 3 minutes followed by ramping down to 25°C at -0.1°C/second. After annealing, 0.8 µL of Klenow polymerase (exonuclease negative, NEB) and 1.35 µL of 1 mM dNTPS was then combined with the 40 µL of product to fill in the barcode oligo (cycling conditions: 25°C for 15:00, 70°C for 20:00, ramp down to 37°C at -0.1°C/s). Digested vector and barcode oligo were then ligated overnight at 16°C.

The overnight ligation was then cleaned and concentrated and eluted in 6 µL of ddH2O. 1 µL of this ligation was then transformed into high efficiency *E. coli* using electroporation at 2 kV. Each reaction contained 1 µL of ligation (or ligation control or pUC19 10 pg/uL) and 25 µL of *E. coli*. 975 µL of pre-warmed SOC media was added to each cuvette after electroporation,

transferred to a culture tube, and recovered at 37°C, shaking for 1 hour. At 1 hour, 1 and 10 µL samples from water and pUC19 cultures were taken and plated on LB supplemented with ampicillin. For ligation and ligation control, four flasks were prepared with 50 mLs of LB and ampicillin, and then 500 µL, 250 µL, 125 µL, 62.5 µL was sampled from the 1 mL of recovery and transferred into a corresponding flask. From those flasks, 1 µL, 10 µL, and 100 µL, were sampled and plated onto LB ampicillin plates. Plates and 50 mL culture were incubated at 37°C overnight. Colonies on plates were then counted, and counts were used to calculate how many unique molecules were transformed to gauge the number of barcodes. Flask with the target number of barcodes was then spun down and midiprepped.

PacBio sequencing and subassembly of CYP2C9 libraries. PacBio libraries were generated using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences) according to manufacturer's directions with the following modifications. Barcoded variant sequences were excised using SpeI-HF and PspXI (activity library) or NheI and SmaI (abundance library) restriction enzymes and purified using AMPure PB beads (Pacific Biosciences 100-265-900) at 1:1 ratio (beads:DNA). Following end-repair and blunt end adaptor ligation, according to manufacturer's instructions, PacBio libraries were subject to 2 additional rounds of restriction digestion to remove any backbone plasmid contamination present in the library. Finally, libraries were cleaned in 3 consecutive rounds of AMPure PB beads (Pacific Biosciences 100-265-900) at 0.6:1 ratio (beads:DNA). The purity and size of Pacbio libraries were confirmed by Tapestation (Agilent) and Bioanalyzer 2100 (Agilent) before proceeding with the sequencing run. Samples were submitted to University of Washington PacBio Sequencing Services and sequenced on two SMRT cells per library in a Sequel run.

Long reads were filtered for at least 10 passes, and analyzed using a custom analysis pipeline to identify gene and barcode regions (https://github.com/shendurelab/AssemblyByPacBio). From PacBio subassembly, the activity library contained 66,958 unique nucleotide variants (22,421 of these full-length), tagged by 105,372 unique barcodes, while the abundance library contained 37,758 unique nucleotide variants (22,669 of these full-length), tagged by 78,740 unique barcodes (Appendix Table A.2).

Tienilic Acid Hexynyl Amide (TAHA) synthesis (activity-based probe). Tienilic Acid (50 mg, 0.15 mmol), EDC (36 mg, 0.18 mmol) and 1-hydroxybenzotriazole hydrate (25 mg, 0.18 mmol), stirring under a nitrogen atmosphere at room temperature, were dissolved in 1 mL of anhydrous acetonitrile and 0.5 mL of anhydrous N,N-dimethylformamide. N-Methylmorpholine (56 μL, 0.45 mmol) was added and the reaction was stirred 15 minutes prior to the addition of hex-5-yn-1-amine (27 μL, 0.18 mmol). The reaction was then stirred another 4 hours after which it was diluted with ethyl acetate and successively washed with 10 % saturated sodium bicarbonate, water and brine. The organic phase was dried over $MgSO_4$ and solvent was evaporated. The final product was purified by flash chromatography, using a hexane/ethyl acetate gradient, and was obtained as a clear oil (52 mg, 84 % yield). ^{1}H NMR (500 MHz, CD_3OD): δ 8.00 (d, J = 4.40 Hz, 1H), 7.48 (d, J = 4.40 Hz, 1H), 7.46 (d, J = 8.79 Hz, 1H), 7.21 (t, J = 4.40 Hz, 1H), 7.14 (d, J = 8.79 Hz, 1H), 4.74 (s, 2H), 3.35 (t, J = 6.83 Hz, 2H), 2.26-2.21 (m, 3H), 1.70 (quin, J = 6.83 Hz, 2H), 1.56 (quin, J = 6.83, 2H).

Flow cytometry-based CYP2C9 functional assay in yeast. CYP2C9 enzymatic activity was probed using a flow-cytometry based method with a click chemistry compatible probe TAHA-ABP (synthesis described above) that has specificity for CYP2C9 activity with minimum

reactivity towards other yeast proteins. Yeast cultures were grown as described above to induce CYP expression, and for each sample, 1 OD of overnight yeast culture was collected via centrifugation at 4000rpm for 2 min, washed with 0.5 mL of PBS by resuspension and centrifugation, and resuspended in 100 µL PBS:0.1% saponin (w/v). Each sample was pre-incubated with 2mM NADPH (Sigma N1630) at 37°C for 20 mins. All samples except a 'No probe' control were treated with 10µM TAHA-ABP and incubated with rotation at 37°C for 20hrs to form activity-dependent CYP2C9-probe adducts. Samples were collected via centrifugation at 4000rpm for 2 min and washed three times with 0.5 mL PBS. Samples were resuspended in 100µL PBS:0.1% saponin (w/v) and incubated at room temperature for 20 mins. Cells were washed twice via resuspension in 1mL PBS and centrifugation at 10,000g for 5 minutes. Cells were resuspended in 100µL copper-catalyzed azide-alkyne cycloaddition (CuAAC) reaction buffer to append a fluorophore reporter (2x concentrations: 10 µM CF488A picolyl azide (Biotium #92187), 2 mM $CuSO_4$ (Sigma C8027), 4 mM THPTA (Sigma 762342), 6 mM ascorbic acid (Sigma A7631) in PBS) and vortexed vigorously to mix. Samples were incubated in the dark at room temperature for 30 minutes and collected by centrifugation as above. Cells were washed five times by resuspension/centrifugation in 0.5 mL PBS, resuspended in 1 mL PBS, and stored at 4°C up to 1 day.

FACS-based deep mutational scan of yeast library. Isogenic humanized yeast strains expressing control *CYP2C9* variants (wildtype, R144C, I359L, and C435H) were induced in galactose as described above. The barcoded *CYP2C9* variant library was thawed at room temperature and ~8 OD of library was inoculated into 25mL YPD media supplemented with 200 µg/ml G418 and grown overnight at 150rpm. The rest of the induction was performed as

described above, with 5x culture volumes and shaking instead of rotation. For each control variant, one sample was collected (1 OD), and for the library, 4 samples (1 OD each) were collected. A “no probe” sample was included as a control. All samples were labeled using the *CYP2C9* functional assay described above with *CYP2C9*-specific activity-based probe TAHA-ABP.

Labeled cells were analyzed using a BD FACSAria III and forward and side scatter (488 nm CHECK), FITC/AF488A (488 nm excitation; 530/30 nm detection filter) parameters were collected. Gates were drawn to contain 10%, 10%, 20%, and 60% of events from the library sample, from most fluorescent (FITC channel) to least fluorescent. Flow cytometry data was collected using FACSDiva (BD Biosciences) and analyzed using FlowJo (Ashland, OR). Four biological replicates of the FACS-based deep mutational scan were performed (Appendix Table A.1).

FACS-based deep mutational scan of VAMP-seq library. HEK293T cells with a serine integrase landing pad integrated via lentivirus with a selectable inducible Caspase 9 cassette (Matreyek et al. 2020) were used for all human cell experiments, enabling expression of a single variant per cell. To recombine variants into HEK293T cells, cells were transfected in 10 cm plates, 3,500,000 cells per plate (4 plates per replicate). 7.1 ug of library plasmid was mixed with 0.48 ug of Bxb1 plasmid in 710 µL of OptiMEM. In a separate tube, 28.5 µL of Fugene was diluted in 685 µL of OptiMEM. The tubes were then combined and incubated at room temperature for 15 minutes. After incubation period, Fugene/DNA mixture was added to cells dropwise, and plates were placed in incubator at 37°C. A minimum of 48 hours after transfection, cells were induced with doxycycline at a final concentration of 2.5 ug/mL. 24 hours

after induction with doxycycline, small molecule AP1903 was added to select from recombinant cells, which causes inducible Caspase 9 in unrecombined landing pads to dimerize and activate.

Recombined HEK293T cells were run on a BD Aria sorter. Cells were gated for live, recombined singlets. For this population, a ratio of eGFP/mCherry was calculated, and the histogram of this ratio was divided into four quartiles. Each quartile was sorted into a 5 mL tube. Sorted cells were grown out for 2-4 days post sorting to ensure enough DNA for sequencing. Three biological replicates of the FACS-based deep mutational scan were performed (Appendix Table A.1)

Sorted library amplification and sequencing. For the yeast activity library, sorted samples were harvested by centrifugation and stored at -20°C. Plasmids were extracted from sorted cell pellets using the Zymoprep Yeast Plasmid Miniprep I kit (Zymo Research D2001). Each sorted sample was split into two for PCR replicates. For each sample, the barcode region was amplified and an 18bp unique molecular identifier (UMI) sequence was added using primers CJA120/CJA124 using KAPA2G Robust HotStart ReadyMix with the following conditions: 95°C for 3m, 2 cycles of 95°C for 20s, 60°C for 15s, 72°C for 30s, and 72°C for 1m, then purified using AMPure XP beads (Beckman Coulter A63880) at 1:1 ratio (beads:DNA). Purified products were amplified using various forward and reverse indexing primers using KAPA2G Robust HotStart ReadyMix with 0.5x SYBR green (Roche #04707516001) on a miniOpticon (Bio-Rad) with the following PCR conditions: 95°C for 3m, up to 30 cycles of 95°C for 20s, 65°C for 15s, 72°C for 30s, and removed from the thermocycler when the relative fluorescence units (RFU) was between 0.5 and 1. These products were again purified using AMPure XP beads (Beckman Coulter A63880) at 1:1 ratio (beads:DNA), and were then gel extracted using the

QIAquick Gel Extraction Kit (Qiagen) and quantified by Qubit fluorometry (Life Technologies). Samples were pooled at equimolar ratios and deep sequenced on an Illumina NextSeq500. Within each sort there was a good correlation of barcode frequencies from PCR replicates (mean Pearson's r=0.859, mean Spearman's ρ=0.694, see Appendix **Figure A.2**).

For the VAMP-seq library, cells were then collected, pelleted by centrifugation and stored at −20 °C. Genomic DNA was prepared using a DNEasy kit, according to the manufacturer's instructions (Qiagen), with the addition of a 30 min incubation at 37 °C with RNAse in the re-suspension step. Eight 50 µl first-round PCR reactions were each prepared with a final concentration of ~50 ng µl−1 input genomic DNA, 1 × Q5 High-Fidelity Master Mix and 0.25 µM of the KAM499/VKORampR 1.1 primers. The reaction conditions were 98 °C for 30 s, 98°C for 10 s, 65°C for 20 s, 72°C for 60 s, repeat 5 times, 72°C for 2 min, 4°C hold. Eight 50 µl reactions were combined, bound to AMPure XP (Beckman Coulter), cleaned and eluted with 21 µl water. Forty percent of the eluted volume was mixed with Q5 High-Fidelity Master Mix; VKOR_indexF_1.1 and one of the indexed reverse primers, PTEN_seq_R1a through PTEN_seq_R2a, were added at 0.25 µM each. These reactions were run with Sybr Green I on a BioRad MiniOpticon; reactions were denatured for 3 minutes at 95°C and cycled 20 times at 95°C for 15s, 60°C for 15s, 72°C for 15s with a final 3 min extension at 72°C. The indexed amplicons were mixed based in relative fluorescence units and run on a 1% agarose gel with Sybr Safe and gel extracted using a freeze and squeeze column (Bio-Rad). The product was quantified using Kapa Illumina quant kit.

Library sequence analysis. For the yeast library, barcode and UMI sequences were trimmed and filtered for minimum base quality Q20 using FASTX-toolkit

(http://hannonlab.cshl.edu/fastx_toolkit/). Barcodes were collapsed according to UMIs using a custom script. UMI-collapsed barcodes were then counted with Enrich2 (Rubin et al. 2017). Barcodes assigned to variants containing insertion, deletions, or multiple amino-acid alterations were removed from the analysis. Custom scripts were used to collapse barcode counts into variant counts. Variants were kept if they had a total (across bin) frequency greater than 1e-5 in each replicate (Appendix Figure A.5). For each replicate, a weighted average of variant frequency across bins was used to determine activity score, with the following bin weights: $w_1 = 0.05$ (bin1), $w_2 = 0.2$, $w_3 = 0.25$, $w_4 = 1$ (bin4). Scores were normalized to the median synonymous weighted average (set to a score of 1), and the median nonsense weighted average of nonsense variants in the first 90% of the protein (score set to 0), and scores were average across replicates. These bin weights were determined by performing a linear regression on activity score vs. normalized geometric mean of fluorescent signal of individual variants using the same activity assay, performed with 14 different alleles spanning the full range of labeling, and taking the bin weights with the best fit (See Supplementary Figure 6). Variants with less than two replicates were removed. Scores for missense variants range from -0.046 to 1.305 and have a bimodal distribution with peaks approximately matching the synonymous and nonsense distributions.

For the VAMP-seq library, barcode sequences were trimmed and filtered using a minimum base quality filter of 20. Barcodes were counted with Enrich2, and barcode counts were collapsed into variant counts as with the yeast library. As before, barcodes assigned to variants containing insertion, deletions, or multiple amino-acid alterations were removed from the analysis. Variants were kept if they had a total (across bin) frequency greater than 1e-4 in

each replicate (Appendix Figure A.5). Abundance scores were calculated as above, but with the following weights: $w_1 = 0.25$ (bin1), $w_2 = 0.5$, $w_3 = 0.75$, $w_4 = 1$ (bin4). Scores were normalized to the synonymous and nonsense distributions as above, but only normalizing to nonsense scores in the middle 80% of positions, excluding the first and last 10% of the protein. Variants with less than two replicates were removed. Scores for missense variants range from -0.29 to 1.59 and have a trimodal distribution with upper and lower peaks approximately matching the synonymous and nonsense distributions.

Activity, abundance, and functional classes were determined as follows, based on a method modified from Matreyek et al. (2018). A synonymous score threshold was used to discriminate between 'wt-like' and 'decreased' scores, this threshold was set at the 5th percentile of synonymous scores (0.879 for activity score, 0.77 for abundance score, and 0.736 for functional score) this. Variants were classified as 'wt-like' if their score and lower confidence interval was greater than the synonymous threshold, or 'possibly wt-like' if just their score was greater than the threshold. Variants were classified as 'decreased' if their score and upper confidence interval was less than the synonymous threshold, or 'possibly decreased' if just their score was less than the threshold. A nonsense score threshold was used to discriminar between 'decreased' and 'nonsense-like' scores, this threshold was the 95th percentile of nonsense scores (0.093 for activity score and 0.282 for abundance score). Variants were classified as 'nonsense-like' if their score and upper confidence interval was less than the nonsense threshold, or 'possibly nonsense-like if just their score was less than the threshold. Finally an upper synonymous threshold was used to discriminate between 'wt-like' and 'increased' score, set at the 95th percentile of synonymous scores (1.102 for activity score and 1.212 for abundance

score). Scores were classified as ‘increased’ if their score and lower confidence interval was greater than the upper synonymous threshold.

We assigned the following activity classes for missense variants (Appendix Figure A.4): nonsense-like (with 568 or 9% of variants), possibly nonsense-like (370 or 6% of variants), decreased (3,049 or 50% of variants), possibly decreased (616 or 10% of variants), possibly wt-like (640 or 10% of variants), wt-like (659 or 11% of variants), and increased activity (240 or 4% of variants). We assigned the following missense abundance classes: nonsense-like (534 or 8% of variants), possibly nonsense-like (377 or 6% of variants), decreased (1,436 or 23% of variants), possibly decreased (823 or 13% of variants), possibly wt-like (1,391 or 22% of variants), wt-like (1,729 or 27% of variants), and increased abundance (80 or 1% of variants).

Small-scale CYP2C9 variant generation and ABPP validation. Individual variants were generated using an inverse PCR-based site-directed mutagenesis approach. Using oligonucleotide pairs whereby the forward primer contains the missense mutation of interest at the 5’ end of the sequence (Integrated DNA Technologies), point mutations were generated using a KAPA HiFi DNA Polymerase (KAPA Biosytems KK2601) and 500pg of *CYP2C9* template sequence p41K*GAL1pr::hCYP2C9-HA*. After performing inverse PCR for each variant, products were run on a 0.7% agarose gel, gel extracted using the QIAquick Gel Extraction Kit (Qiagen), treated with T4 polynucleotide kinase (NEB M0201) at 37 °C for 30min, and ligated with T4 DNA ligase (NEB M0202) at 16 °C overnight. Ligated products were used to transform chemically competent *E. coli* cells (NEB C2987 or Bioline BIO-85027). Bacterial clones were prepared for plasmid extraction using the QIAprep Spin Miniprep Kit (Qiagen) and variant sequences were confirmed with Sanger sequencing. Plasmids containing missense variants were

individually transformed into YMD4256 using the 1-step transformation protocol (Chen et al. 1992) and selection for growth in YPD supplemented with 200 μg/ml G418. Individual clones from each transformation were stored at -80 °C.

Individual CYP2C9 yeast-expressed variants were grown and induced in galactose as described above, and 1 OD/ml of culture was collected for each variant. All samples were labeled using the *CYP2C9* functional assay described above with *CYP2C9*-specific activity-based probe TAHA-ABP. Labeled cells were analyzed using a BD LSRII and forward and side scatter (488 nm CHECK), FITC (488 nm excitation; 530/30 nm detection filter) parameters were collected for 20,000 events. Flow cytometry data was collected using FACSDiva (BD Biosciences) and analyzed using FlowJo (Ashland, OR). Fluorescence (FITC geometric mean of gated single cells) was normalized to background labeling ('no probe' control) and variant ABPP labeling relative to WT was calculated. Three biological replicates of CYP2C9 individual variant validation were performed.

Yeast microsomal preparations. A large-scale induction of yeast cells expressing *CYP2C9* or variant plasmid was done as described above with a final culture volume of 0.5 L. After switching cells to galactose culture, cells were collected after 18-22hrs. Cells were pelleted and stored at -80°C until ready for microsome preparations.

Yeast microsomes were prepared as described previously (Pompon et al. 1996, McDonald et al. 2017) with slight modifications. Harvested cells were thawed at room temperature for at least 10 minutes, washed with 25 ml of TEK buffer (50 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1 MKCl), recovered at 3200g, resuspended in 30 ml of TEM buffer (50 mM Tris-HCl, pH 7.4, 1 mMEDTA, 70 mM 2-mercaptoethanol), and incubated at room temperature

for 10 minutes. Cells were recovered (3200g) and resuspended in 1.5 ml of TMS buffer (1.5 M sorbitol; 20 mM Tris-MES, pH 6.3; 2 mM EDTA), and 20 mg of 20T Zymolyase was added. Cells were incubated for 1-1.15 hour at 30°C with agitation until digested. Further steps were performed on ice. Spheroplasts were pelleted at 6732g and washed with 25 ml of TES-A buffer (50 mM Tris-HCl, pH 7.4, 1 mM EDTA, 1.5 M sorbitol), and the centrifugation step was repeated. Spheroplasts were resuspended in 10 ml of TES-B buffer (50 mM Tris-HCl, pH 7.4; 1 mM EDTA; 0.6 M sorbitol) and lysed using a Misonix S4000 by performing 4 x 15 second pulses at maximum amplitude (40-45W). After 5 minutes on ice, lysed cells were centrifuged for 4 minutes at 1700g. The supernatant was then centrifuged at 110,000g for 70 minutes. The microsomal pellet was resuspended in 1 ml of TEG buffer (50 mM Tris-HCl pH 7.4, 1 mM EDTA, 20% (v/v) glycerol), homogenized, and frozen at -80°C.

Warfarin metabolism validation assay. S-Warfarin (50 μM) was mixed together with yeast lysate, prepared from CYP2C9 variant-expressing cells, at 5 mg/mL total protein in 100 mM KPi buffer, pH 7.4 (100 μL final incubation volume). After 3 minutes pre-incubation at 37 ^{o}C in a water bath, NADPH was added to initiate (to 1 mM final concentration). Reactions were incubated for 20 minutes and were then quenched with the addition of 5 μL of ice-cold 70% $HClO_4$. An internal standard solution, containing 5 ng each of 6-hydroxywarfarin-d_5 and 7-hydroxywarfarin-d_5, was added and the reaction products were vortexed and centrifuged to remove protein. Supernatants were analyzed by LC-MS/MS. Three technical replicates were carried out for each CYP2C9 variant lysate. Calibration curves were prepared by spiking variable amounts of unlabeled 6- and 7-hydroxywarfarins into 100 μL volumes of KPi buffer in order to generate standard mixtures with final concentrations ranging from 1 nM to 1 μM. These

standard solutions were worked up and analyzed in an identical fashion to that described for the incubation samples.

Phenytoin metabolism validation assay. Phenytoin (100 μM) was mixed together with yeast lysate, prepared from CYP2C9 variant-expressing cells, at 5 mg/mL total protein in 100 mM KPi buffer, pH 7.4 (200 μL final incubation volume). After 3 minutes pre-incubation at 37 ^{o}C in a water bath, NADPH stock (to 1 mM final concentration) was added to initiate the reactions. Reactions were incubated for 20 minutes and were then quenched with the addition of 20 μL of ice-cold 15% $ZnSO_4$. 4-Hydroxyphenytoin-d_5 (p-HPPH-d_5, 10 ng) was added as the internal standard and the reactions were vortexed, then centrifuged to remove protein, and the supernatants were analyzed by LC-MS/MS. Again, three technical replicates were carried out for each CYP2C9 variant lysate. Calibration curves were prepared by spiking variable amounts of unlabeled 4-hydroxyphenytoin (p-HPPH) into 200 μL volumes of KPi buffer, generating standard mixtures with final concentrations ranging from 1 nM to 1 μM. These standard solutions were worked up and analyzed in an identical fashion to that described for the incubation samples.

LC-MS/MS of Warfarin and Phenytoin Metabolites. LC-MS/MS analyses of warfarin and phenytoin metabolic reactions were conducted on a Waters Xevo TQ-S Tandem Quadrupole Mass Spectrometer (Waters Co., Milford, MA) coupled to an ACQUITY Ultra Performance LC™ (UPLC™) System with integral autoinjector (Waters). The Xevo was operated in ESI^{+}-MS/MS (SRM) mode at a source temperature of 150^{o}C and a desolvation temperature of 350^{o}C. The following mass transitions were monitored in separate ion channels for the various oxidative warfarin metabolites/standards: *m/z* 325 > 179 (6- and 7-hydroxywarfarins-d_0) and *m/z*

330 > 179 (6- and 7-hydroxywarfarins-d_5); and phenytoin metabolite and standard: *m/z* 269 > 198 (p-HPPH-d_0) and *m/z* 274 > 203 (p-HPPH-d_5). Optimized cone voltages and collision energies were set to 25 V and 15 eV for all metabolites and standards of warfarin, while the cone voltage was set to 35 V with a collision energy of 15 eV for the phenytoin metabolite p-HPPH (both d_0 and d_5-labeled). Metabolic products from the warfarin incubations were separated on an Acquity BEH Phenyl, 1.7 μ, 2.1 x 150 mm UPLC column (Waters, Corp) using an isocratic gradient of 45% solvent A (0.1% aqueous formic acid) and 55% solvent B (methanol), with a constant flow rate of 0.35 mL/min. Phenytoin metabolites were separated using this same BEH Phenyl UPLC column with a solvent gradient of water (solvent A) and acetonitrile (solvent B), both of which contained 0.1% formic acid, running at a flow rate of 0.3 mL/min. Initially, solvent B was set to 28%, where it was maintained for 4.5 minutes, then increased linearly to 95% over 0.5 minutes where it was left for an additional 1.5 minutes. Metabolites were quantified through comparison of their peak area ratios (relative to either the 6- and 7-hydroxywarfarin-d_5 or p-HPPH-d_5 internal standard peak areas) to calibration curves using linear regression analysis. The limits of detection for all of the metabolites were below 5 fmol injected on column. Mass spectral data analyses for the Xevo TQ-S were performed on Windows XP-based Micromass MassLynxNT, v. 4.1, software (Waters).

BOMCC fluorogenic assay. 7-Benzyloxymethyloxy-3-cyanocoumarin (BOMCC) (50 μM) was mixed together with 200 μM NADPH and yeast lysate at 50 μg total protein, prepared from CYP2C9 variant-expressing cells, in 50 mM KPi buffer, pH 8 (150 μL final incubation volume). Each sample was done in parallel with a no NADPH control. Three technical replicates were carried out for each CYP2C9 variant lysate. Sample fluorescence (excitation: 410nm,

emission: 460nm, gain: 60) was recorded every 5 minutes on a BioTek Synergy H1 microplate reader at 37°C for 200 min with shaking. To determine relative activity, the fluorescence from each sample was normalized by subtracting the no NADPH control, and the slope of the normalized fluorescence signal during the linear range (5 mins to 50 mins) was calculated. Slopes were averaged across technical replicates, and normalized by WT average slope to determine BOMCC metabolism relative to WT.

Chapter 5. Summary and Future Directions

In these chapters, I described my work developing high-throughput methods of studying human cytochrome P450 activity in yeast. In Chapter 2, I discussed methods of improving heterologous CYP expression and activity in yeast through targeted strain engineering, screening of natural isolates, and experimental evolution. In Chapter 3, I tried several approaches to develop a pooled CYP2C9 activity assay, ultimately developing an assay leveraging activity-based protein profiling with novel CYP2C9 activity-based probes. This assay was used in Chapter 4 to perform a deep mutational scan of CYP2C9, generating functional annotation for thousands of variants. In this chapter, I outline future avenues of study and analyze my work in the larger context of the field of pharmacogenomics.

5.1 Promise of yeast natural isolates for CYP activity assays

The Design-Build-Test-Learn cycle is a common paradigm in industrial and metabolic engineering, used to engineer microbes to produce new compounds at high yields, or to develop new phenotypes. Incorporating modern approaches in the Design-Build-Test-Learn cycle, especially comparative genomics, has been extremely successful in industrial engineering (Sardi

and Gasch 2017). Adaptive laboratory evolution is also a key approach to modern industrial engineering (Winkler and Kao 2014, Shepelin et al. 2018). In Chapter 2, we took several different approaches to engineer yeast to be an optimal CYP expression system, including targeting strain modification, natural isolates, and experimental evolution, undergoing several rounds of the Design-Build-Test-Learn cycle. Using a diversity of approaches was necessary since we did not fully understand the genetics behind our target phenotype. Ultimately, a combination of screening natural isolates and targeted strain humanization produced a much improved strain.

Additionally, we only screened 11 different yeast natural isolates from the SGRP collection (Liti et al. 2009), and a much larger (1,011 strains) collection of yeast natural isolates is now available (Peter et al. 2018). In the future, screening this collection for human CYP expression and activity would likely result in many candidate strains with favorable phenotypes. This would allow us to cherry-pick a candidate strain that lacked unfavorable phenotypes such as cell aggregation and flocculation. A portion of the SGRP collection has been barcoded (Cubillos et al. 2009), so we could use this collection to screen innate CYP expression *en masse* by using Click-seq on a library of barcoded strains, and could even compare across multiple human CYP isoforms. During our small-scale screening of natural isolates, we found that sake strain Y12 produced high levels of active human CYP enzyme, so it would be interesting to see whether strains of the same ecological origin or niche perform similarly.

5.2 Challenges of developing high-throughput cell-based assays of CYP activity

One major bottleneck to performing deep mutational scans is library generation, especially with larger proteins, where it is challenging to produce a complete library containing

close to 100% of all possible single mutations. With our CYP2C9 libraries (490 amino acids), we were only able to generate 67% coverage for the yeast activity and 85% coverage for the human abundance libraries (percent possible single amino acid mutations), and ultimately produced activity and abundance scores for 67% and 70% of all possible single amino acid mutations. Separate issues resulted in incomplete coverage for each library.

Some variant loss likely occurred during the amplification and cloning process, but differences in PacBio long read coverage during subassembly led to large reductions in starting library coverage for the yeast activity library. The activity library (with less reads and more barcodes) had 2.8x read coverage compared to 6.9x coverage of the abundance library (with more reads and less barcodes). This led to a higher fraction of barcodes associated with indels due to PacBio sequencing error in the activity versus the abundance library. In the activity library, 50% of barcodes were associated with an indel while only 23% of barcodes in the abundance library were associated with an indel. With libraries in the future, ensuring adequate coverage will likely improve this issue, although we could also be dealing with underlying library quality issues unrelated to PacBio sequencing. Though the starting abundance library was more complete than the activity library, we observed further variant loss in the abundance library during VAMP-seq, which may have been due to low replicate numbers or low numbers of cells sorted. Conversely, with the activity assay, almost all variants in the starting library were retained after Click-seq sequencing. Ultimately, a wide variety of challenges during library generation can result in missing functional data for some percentage of variants. One possible solution for this is data imputation, which has been successfully applied to some DMS datasets (Weile et al. 2017).

Several limitations specific to Click-seq are cell inviability, slow labeling incubation period, and lack of expression controls. We have discussed the first two limitations in Chapter 3. To improve Click-seq, we would like to normalize activity scores to CYP expression levels within the yeast system itself, rather than comparing to VAMP-seq results generated in a different experimental system. There are several approaches we could take to do this. First, we could use immunofluorescence to directly quantify CYP protein levels. Since our CYP2C9 library was HA-tagged, we can use a fluor-conjugated HA antibody to label our library, followed by FACS. Click-seq and immunofluorescence could be performed on subsets of the same biological replicate population, adding to the validity of this approach. Alternatively, RNA-seq could be performed on ABPP-labeled and sorted yeast cells, since variant barcodes are located within the 3' UTR. Although this would be a measure of transcript expression rather than a direct quantification of protein expression, this would be a relatively simple method to employ as a readout of CYP transcript levels.

5.3 Extension of P450 deep mutational scan to multi-variant libraries

Depending on the mutagenesis method, variant libraries can be limited to single codon variants or have a range of variant types and combinations (Weile and Roth 2018). Our CYP2C9 libraries were generated using inverse PCR with primer pairs targeted to each codon, with the goal of generating all possible single codon changes. This approach guarantees that the library is mostly single mutations, but due to PCR and other errors a small fraction of the library will invariably be made up of multiple mutations. 9.4% of the CYP2C9 activity library and 9.6% of the CYP2C9 abundance library was composed of alleles with two or more amino acid mutations. We would like to investigate variants with two amino acid mutations for which the

corresponding single amino acid mutations also exist in the library (this may be a small number of variants), and determine whether any of these variants display non-additive effects.

CYP2C9 *18, *35, and *61 alleles (https://www.pharmvar.org) have multiple missense variants, with common variant R144C (*2) or I359L (*3) co-occurring with a second (rare) missense allele. These multi-variant alleles have not been well characterized, so to investigate larger epistatic trends, we could use site directed mutagenesis to add common alleles to the background of our variant library, resulting in new libraries with multiple mutations. It would be interesting to add common variants *2 (R144C), *3 (I359L), *8 (R150H), and *9 (H251R) to our CYP2C9 activity library, and for each common variant use Click-seq to characterize variant function. Though only three star alleles with multiple amino acid variants have been documented in CYP2C9, the incidence of rare/common variants on the same haplotype is likely quite frequent due to *de novo* variation and recombination, and characterization of non-additive effects is important.

There are some mutation types that our yeast cDNA expression system is not applicable for. First, since the CYP2C9 library was codon-optimized, we cannot extrapolate any synonymous variant effects to human P450 expression differences. Secondly, we cannot measure any splice variants or non-coding variants, even though these mutations have documented effects on CYP function. The *CYP2C19* *17 promoter variant (which is quite common) leads to increased activity (Sim et al. 2006), while the *CYP2C9* *8 R150H mutation shows strong linkage disequilibrium with two promoter polymorphisms that result in lower levels of transcription (Cavallari et al. 2013). Finally, our library design does not allow us to measure the effect of larger structural variants.

5.4 Extension of the Click-seq approach to other CYP2C9 substrates

Though we were able to show that our Click-seq approach with CYP2C9 using a tienilic acid-derived probe had results generalizable to other substrates in small scale validation (warfarin and phenytoin), we may be missing substrate-dependent effects. These have been documented with *2 and *3 alleles (Lee et al. 2002). In general, drug metabolism CYPs have a broad specificity and can bind many different substrates, though this varies by isoform. CYP2C9 has evidence for multiple ligand binding sites (Maekawa et al. 2017) and has a large active site cavity (Porubsky et al. 2008).

To investigate this, we would like to repeat the Click-seq approach with additional CYP2C9 substrates to determine substrate-dependent effects. Although the warfarin and cyclocoumarol-derived activity based probes we tested were ineffective, we had success with a novel tolbutamide-derived probe and a general P450 dialkyne probe ABP5 (Wright et al. 2009). The tolbutamide-derived probe would need further optimization to improve labeling, while ABP5 did not distinguish between intermediate activity alleles very well but had a strong labeling signal. If these probes do not show enough promise, we could develop new probes from other CYP2C9 mechanism based inhibitors. In addition to teinilic acid, known mechanism based inhibitors of CYP2C9 include suprofen (O'Donnell et al. 2003) and silybin (Sridar et al. 2004) and could be used to design new probes.

Additionally, measuring competitive substrate inhibition is an alternative approach to measure substrate-specific effects. In this approach, Click-seq can be coupled with a competitive inhibitor, where addition of a CYP2C9 substrate will prevent TAHA activity-based probe binding. When tested across a range of substrate concentrations, this approach can be used to

quantify substrate-dependent variant activity for the entire CYP2C9 library. This has been shown to work with quinidine and dextromethorphan for CYP2D6 (unpublished results). To implement this for CYP2C9, we would perform pilot studies with an allele series with substrates (S)-warfarin, phenytoin, and flurbiprofen to test for competitive inhibition.

5.5 Extension of the Click-seq approach to other CYP enzymes

Click-seq, the method employed in Chapter 4 to measure CYP2C9 variant activity *en masse*, can be extended to other CYP enzymes. Of the suite of general P450 probes developed by Wright et al. (2009), we found that several of these probes, ABP5 and ABP6, label CYP2D6 well. We expect similar success with other CYPs, though expression level differences may have a large effect on the successfulness of labeling in yeast. Additionally, the original set of CYP activity-based probes are hydrophobic dialkynes that are reactive towards many P450s, rather than selective for a specific CYP. However, when developing probes for CYP2C9, we saw better labeling with a probe derived from a CYP2C9-specific inhibitor. Thus for other CYPs, we expect more success with novel activity-based probes derived from CYP covalent inhibitors and other substrates. For CYP2D6, we have already generated and tested a novel activity-based probe derived from ticlopidine, a covalent inhibitor of several CYPs , including CYP2D6 and CYP2C19 (Ko et al. 2000). This probe was effective at labeling CYP2D6, but has not yet been evaluated with CYP2C19.

In addition to CYP2C9, CYP2C19, and CYP2D6, several other CYPs have have been assigned level A or B priority in a gene-drug pair by CPIC, indicating that CYP genotype can or should be used to change dosage of the drug in question (https://www.pharmgkb.org/cpic/pairs). CPIC has designated CYP2B6 (with efavirenz, methadone, nevirapine), CYP3A5 (tacrolimus),

and CYP4F2 (phenprocoumon, warfarin, acenocoumarol) high priority. Additionally, CYP2A6 is involved in nicotine metabolism and relevant for drug addiction (Chenoweth et al. 2014), and CYP3A4 has the largest substrate specificity of the human P450s and can metabolise many substrates. Any of these CYPs would be important to profile with Click-seq. We could screen these CYPs against the probes developed by Wright et al. (2009) since many of those probes show reactivity with many CYPs, especially the 2-ethynylnaphthalene-derived probe.

Additionally, we would like to also screen these CYPs with VAMP-seq, as generating both an activity and abundance dataset for a given CYP allows for more nuanced interpretation of variant function. With CYP2C9, we generated two separate variant libraries, one codon-optimized for yeast and one codon-optimized for human expression. Though codon-optimization is commonly used to improve heterologous expression in yeast systems (Lanza et al. 2014), this step is not strictly necessary, and many earlier studies expressed human P450 in yeast as cDNAs without any codon optimization (Oeda et al. 1985, Shibata et al. 1990). In the future, we could express both yeast-optimized and human-optimized P450 in yeast to determine whether we could use the same sequence for Click-seq and VAMP-seq, which would allow us to generate one library for two assays. CYP2D6 is the best candidate for a pilot test, since it has the highest expression levels in yeast of the CYPs tested.

Finally, since library generation is an arduous and time intensive step in DMS, it may be infeasible to perform deep mutational scans on more than a handful of CYP genes. In that case, we would use machine learning to determine how well DMS data from multiple CYPs can predict variant function in a different CYP (within and across families and subfamilies). We would use leave-one-out cross validation to evaluate our predictions. We expect highly similar

CYPs (for example, CYP2C9 and CYP2C19, which have a pairwise amino acid identity of 91%) to have similar patterns in functional scores compared to more diverged pairs of CYPs.

5.6 Massively parallel assays with other pharmacogenes

Click-seq can be extended to any CYP enzyme with the right probe, but in order to perform deep mutational scans with non-CYP pharmacogenes, alternate assays specific to individual pharmacogenes will need to be developed. Other important pharmacogenes with high priority (A or B) gene-drug pairs as determined by CPIC include *VKORC1*, *TPMT*, *G6PD*, *DYPD*, *SLCO1B1*, *IFNL3*, and *NUDT15*. Deep mutational scans have already been performed with TMPT, NUDT15, and VKORC1 (Matreyek at al. 2018, Suiter et al. 2020, Chiasson et al. 2020). One general DMS approach is VAMP-seq (Matreyek at al. 2018), which measures protein stability and can be used with any pharmacogene (and was used with *TPMT*, *NUDT15*, and *VKORC1*), but additional assays of variant function or activity are invaluable for clinical utility of DMS datasets.

Genetic variation in glucose-6-phosphate dehydrogenase (*G6PD*) can lead to G6PD deficiency, which affects 400 million people worldwide (Mehta et al. 2000). In order to test G6PD variant function *en masse*, a yeast complementation approach can be used. Human *G6PD* complements yeast ortholog *ZWF1*, the deletion of which leads to methionine auxotrophy. This complementation has been used to characterize a human G6PD allele in yeast (Grabowska et al. 2004), and a similar approach was used to characterize 53 human G6PD variants in yeast under oxidative stress conditions (Merritt et al. 2005). Thus, once a G6PD library is generated, a simple auxotrophic growth selection can be used to determine the function of thousands of G6PD variants.

For other pharmacogenes, determining a suitable growth assay similar to G6PD would be ideal. Systematic humanization of yeast genes showed that 200 of 414 essential yeast genes were complemented by their human ortholog and could rescue a lethal growth defect (Kachroo et al. 2015), but none of the above CPIC high-priority pharmacogenes are included. Determining a growth assay for high priority pharmacogenes in yeast will require further investigation and pilot studies.

5.7 Pharmacogenomics in practice: The utility and limitations of deep mutational scanning datasets

Many groups are involved in the curation and dissemination of pharmacogenomic knowledge for clinical implementation. The Pharmacogene Variant (PharmVar) Consortium is a repository of allelic variants of cytochrome P450s and other pharmacogenes, and is a hub for pharmacogene nomenclature (Gaedigk et al. 2018). The Pharmacogenomic Knowledgebase (PharmGKB) is a vital PGx resource for both researchers and clinicians, and in 2013 had more than 5,000 variant annotations from 900 PGx-related genes associated with more than 900 drugs (Thorn et al. 2013). The Clinical Pharmacogenetics Implementation Consortium (CPIC), established jointly between PharmGKB and the Pharmacogenomics Research Network, curates gene/drug recommendations and publishes guidelines for how to implement pharmacogenomic information in a clinical setting (Relling and Klein 2011). As of July 2020, CPIC has published guidelines for 19 genes and 54 drugs (https://www.pharmgkb.org/guidelineAnnotations).

To encourage the implementation of PGx-based dosing in routine clinical care, the Translational Pharmacogenetic Program (TPP) was established in 2011 at eight different healthcare sites (Luzum et al. 2017). At these sites, the gene-drug pairs most commonly

implemented for clinical decision support (CDS) are the following: CYP2C19-clopidogrel, CYP2C9 and VKORC1-warfarin, CYP2D6-codeine, SLCO1B1-simvastatin, and TPMT-thiopurines (Dunnenberger et al. 2015). To implement CDS, patient genotype and predicted metabolizer phenotype are entered into the electronic health record preemptively, and later, when a prescription is ordered for a patient with an actionable phenotype, alerts for point-of-care decision support are deployed (Pulley et al. 2011). As an example, at the Mayo Clinic, a patient with the genotype *CYP2C19 *1/*2* (normal/no activity diplotype) would be classified as an "intermediate metabolizer" (as opposed to ultrarapid, extensive, or poor metabolizer), and a clinician would be prompted to consider an alternative therapy before prescribing clopidogrel. This decision support was implemented with 11 *CYP2C19* alleles and alleles from four other pharmacogenes (Ji et al. 2016).

Pharmacogenomic implementation suffers from many barriers, including insufficient PGx functional annotation and clinical guidelines, the challenge of upgrading electronic health systems to report and store genetic test results, clinicians' hesitance to use genetic data, and the costs of genotyping and lack of insurance reimbursement. There are conflicting results on whether PGx implementation is cost effective or even effective in improving clinical care. One notable example is with warfarin, where genotype-based dosing was tested in two large randomized controlled trials (Pirmohamed et al. 2013, Kimmel et al. 2013) using the common variants *VKORC1* -1639G>A and *CYP2C9*2* and *3. Only the Pirmohamad et al. study found that genotype-based dosing outperformed standard dosing practices. This may have been due to differences in ethnic heterogeneity, as the Pirmohamed et al. study was 99% Caucasian, while the Kimmel et al. study was 73% Caucaian and 27% African American (Pirmohamed et al.

2015). *CYP2C9*2* and **3* allele frequencies are much lower in Africans than in Europeans (12% and 6% compared to 2% and 1%) (Zhou et al. 2017), and other *CYP2C9* alleles common in Africans such as *CYP2C9*8* and **9* were not included in dosing guidelines. Including more variants in clinical dosing guidelines should improve patient care across populations.

Implementing data from multiplexed assays of variant effects (MAVEs) and deep mutational scanning into clinical settings is relatively uncharted territory. The American College of Medical Genetics and Genomics (ACMG) has developed guidelines for the interpretation of sequence variants in genes that cause Mendelian disorders. This includes guidance for how functional data should be used to lend varying levels of support towards a benign or pathogenic interpretation (Richards et al. 2015). Well-established *in vivo* or *in vitro* functional studies can be used as varying levels of evidence towards a benign or pathogenic designation, which includes data from MAVEs/DMSs, although the ACMG guidelines were originally intended for traditional low-throughput assays. More recent recommendations have begun to clarify how to evaluate functional assays, such as requiring a minimum of 11 total pathogenic and benign variant controls to reach moderate-level evidence (Brnich et al. 2019). Additionally, recommendations have been put forward for how best to use data from multiplexed functional assays for ACMG variant interpretation (Gelman et al. 2019). Moreover, data from recent high-throughput functional assays of BRCA1 (Findlay et al. 2018, Starita et al. 2015) has been used to reclassify variants of uncertain significance (VUSs) according to ACMG guidelines (Kim et al. 2020). However, determining variant function in pharmacogenes is additionally difficult because phenotypes are only apparent upon exposure to a drug.

With any sort of deep mutational scan, it is necessary to have enough variants with curated clinical evidence of "normal function" or "no function" (or benign/pathogenic) to determine assay predictive value, commonly at least 10 - 20 per category. However, when evaluating our CYP2C9 DMS activity data, a dearth of alleles with 'normal function" designations by CPIC (Relling and Klein 2011) meant that it was difficult to determine a predictive value, even though our assay correlated well with gold-standard low-throughput assays. This is somewhat specific to *CYP2C9*, as *CYP2C19* and *CYP2D6* each have >5 documented alleles of normal function (www.pharmvar.org). To solve this issue, our DMS data has identified several *CYP2C9* star alleles that could be used as normal function reference alleles in the future, provided those alleles undergo more extensive biochemical characterization.

In clinical settings, CYP allele functional data is collapsed into substrate-agnostic categories such as "normal function", "decreased function", or "no function", and the simplicity of this categorization is meant to help with interpretation. CPIC *CYP2C9* allele dosing guidelines are nearly identical for warfarin (Johnson et al. 2017) and NSAIDs (Theken et al. 2020), even though there are documented substrate-dependent effects with certain *CYP2C9* alleles (Lee et al. 2002). In small-scale validation, our CYP2C9 DMS activity data correlated better with S-warfarin activity than phenytoin activity and may be better suited to inform warfarin dosing, though is likely still informative for both. We would like to perform biochemical studies of individual CYP2C9 variants with more substrates to better understand these substrate-dependent effects, and also to perform Click-seq with a different CYP2C9 activity-based probe if possible. However, we will ultimately only know which substrates our DMS datasets are informative for if we compare them to a clinical dataset or perform a clinical trial. One possibility would be to

compare our predictions to a large warfarin cohort with *CYP2C9* genotype and electronic health record information to determine whether patients with predicted low activity alleles had increased risk of a bleeding event. This approach was taken by Higashi et al. in a retrospective cohort study (2002), though a larger population would be needed to look at alleles beyond just *2 and *3.

Another barrier to clinical DMS implementation is how to report functional data from one or more DMS datasets. With CYP2C9, a reasonable approach is to bin our continuous activity scores into categories according to CPIC classifications (normal, decreased, no function) for ease of interpretation, but we would likely report the continuous activity score as well in order to provide a more nuanced clinical recommendation. Additionally, we have a second DMS dataset of CYP2C9 abundance and can calculate a combined activity-abundance score, as was done with NUDT15 (Suiter et al. 2020). It will be important to determine whether such a CYP2C9 combined functional score is more informative than a single score, and whether we expect variants with only an abundance score to be informative clinically. Moreover, there are currently no repositories for disseminating PGx-specific MAVE data with researchers and clinicians, other than general repository MaveDB (Esposito et al. 2019). Developing PGx-specific terminology and databases for sharing DMS functional data will require the expertise of PharmGKB, CPIC, and other organizations in order to ensure accurate interpretation of variant functional scores by clinicians.

The work described in these chapters contributes to knowledge of pharmacogenomics and P450 biology. We developed a massively parallel yeast-based assay of CYP activity by leveraging activity-based protein profiling and high-throughput sequencing, called Click-seq. We

used this method to generate activity scores for thousands of CYP2C9 variants, and compared them to an orthogonal assay that measured variant abundance at scale. These CYP2C9 datasets are a rich resource of functional annotation for clinical implementation of pharmacogenomics, and also a detailed map of CYP2C9 structure-function. Click-seq and the methods discussed in these chapters can be extended to other human CYP enzymes, and are a general framework for generating PGx variant functional annotation at a massively parallel scale.

Bibliography

Ahler, E., Register, A. C., Chakraborty, S., Fang, L., Dieter, E. M., Sitko, K. A., Vidadala, R. S. R., Trevillian, B. M., Golkowski, M., Gelman, H., Stephany, J. J., Rubin, A. F., Merritt, E. A., Fowler, D. M., & Maly, D. J. (2019). A combined approach reveals a regulatory mechanism coupling src's kinase activity, localization, and phosphotransferase-independent functions. *Molecular Cell*, *74*(2), 393-408.e20. https://doi.org/10.1016/j.molcel.2019.02.003

Aklillu, E., Persson, I., Bertilsson, L., Johansson, I., Rodrigues, F., & Ingelman-Sundberg, M. (1996). Frequent distribution of ultrarapid metabolizers of debrisoquine in an ethiopian population carrying duplicated and multiduplicated functional CYP2D6 alleles. *The Journal of Pharmacology and Experimental Therapeutics*, *278*(1), 441–446.

Arendse, L. B., & Blackburn, J. M. (2018). Effects of polymorphic variation on the thermostability of heterogenous populations of CYP3A4 and CYP2C9 enzymes in solution. *Scientific Reports*, *8*(1), 11876. https://doi.org/10.1038/s41598-018-30195-1

Asseffa, A., Smith, S. J., Nagata, K., Gillette, J., Gelboin, H. V., & Gonzalez, F. J. (1989). Novel exogenous heme-dependent expression of mammalian cytochrome P450 using baculovirus. *Archives of Biochemistry and Biophysics*, *274*(2), 481–490. https://doi.org/10.1016/0003-9861(89)90461-x

Backes, W. L., & Kelley, R. W. (2003). Organization of multiple cytochrome P450s with NADPH-cytochrome P450 reductase in membranes. *Pharmacology & Therapeutics*, *98*(2), 221–233. https://doi.org/10.1016/s0163-7258(03)00031-7

Barnes, H. J., Arlotto, M. P., & Waterman, M. R. (1991). Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(13), 5597–5601.

Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., Nguyen Ba, A. N., Moses, A. M., Louis, E. J., Mustonen, V., Warringer, J., Durbin, R., & Liti, G. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, *31*(4), 872–888. https://doi.org/10.1093/molbev/msu037

Bielinski, S. J., Olson, J. E., Pathak, J., Weinshilboum, R. M., Wang, L., Lyke, K. J., Ryu, E., Targonski, P. V., Van Norstrand, M. D., Hathcock, M. A., Takahashi, P. Y., McCormick, J. B., Johnson, K. J., Maschke, K. J., Rohrer Vitek, C. R., Ellingson, M. S., Wieben, E. D., Farrugia, G., Morrisette, J. A., … Kullo, I. J. (2014). Preemptive genotyping for personalized medicine: Design of the right drug, right dose, right time – using genomic data to individualize treatment protocol. *Mayo Clinic Proceedings*, *89*(1), 25–33. https://doi.org/10.1016/j.mayocp.2013.10.021

Bogengruber, E., Eichberger, T., Briza, P., Dawes, I. W., Breitenbach, M., & Schricker, R. (1998). Sporulation-specific expression of the yeast DIT1/DIT2 promoter is controlled by a newly identified repressor element and the short form of Rim101p. *European Journal of Biochemistry*, *258*(2), 430–436. https://doi.org/10.1046/j.1432-1327.1998.2580430.x

Brnich, S. E., Abou Tayoun, A. N., Couch, F. J., Cutting, G. R., Greenblatt, M. S., Heinen, C. D., Kanavy, D. M., Luo, X., McNulty, S. M., Starita, L. M., Tavtigian, S. V., Wright, M. W., Harrison, S. M., Biesecker, L. G., Berg, J. S., & Clinical Genome Resource Sequence Variant Interpretation Working Group. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Medicine, 12(1), 3. https://doi.org/10.1186/s13073-019-0690-2

Budnitz, D. S., Pollock, D. A., Weidenbach, K. N., Mendelsohn, A. B., Schroeder, T. J., & Annest, J. L. (2006). National surveillance of emergency department visits for outpatient adverse drug events. *JAMA*, *296*(15), 1858–1866. https://doi.org/10.1001/jama.296.15.1858

Caraco, Y., Muszkat, M., & Wood, A. J. (2001). Phenytoin metabolic ratio: A putative marker of CYP2C9 activity in vivo. *Pharmacogenetics*, *11*(7), 587–596. https://doi.org/10.1097/00008571-200110000-00005

Cavallari, L. H., Vaynshteyn, D., Freeman, K. M., Wang, D., Perera, M. A., Takahashi, H., Drozda, K., Patel, S. R., & Jeong, H. (2013). CYP2C9 promoter region single-nucleotide polymorphisms linked to the R150H polymorphism are functional suggesting their role in CYP2C9*8-mediated effects. *Pharmacogenetics and Genomics*, *23*(4), 228–231. https://doi.org/10.1097/FPC.0b013e32835e95c7

Chauret, N., Dobbs, B., Lackman, R. L., Bateman, K., Nicoll-Griffith, D. A., Stresser, D. M., Ackermann, J. M., Turner, S. D., Miller, V. P., & Crespi, C. L. (2001). The use of 3-[2-(N,N-diethyl-N-methylammonium)ethyl]-7-methoxy-4-methylcoumarin (Ammc) as a specific CYP2D6 probe in human liver microsomes. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *29*(9), 1196–1200.

Chen, D. C., Yang, B. C., & Kuo, T. T. (1992). One-step transformation of yeast in stationary phase. *Current Genetics*, *21*(1), 83–84. https://doi.org/10.1007/BF00318659

Chenoweth, M. J., O'Loughlin, J., Sylvestre, M.-P., & Tyndale, R. F. (2013). CYP2A6 slow nicotine metabolism is associated with increased quitting by adolescent smokers. *Pharmacogenetics and Genomics*, *23*(4), 232–235. https://doi.org/10.1097/FPC.0b013e32835f834d

Chiasson, M. A., Rollins, N. J., Stephany, J. J., Sitko, K. A., Matreyek, K. A., Verby, M., Sun, S., Roth, F. P., DeSloover, D., Marks, D., Rettie, A. E., & Fowler, D. M. (2020). Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact. *BioRxiv*, 2020.05.10.087312. https://doi.org/10.1101/2020.05.10.087312

Chiasson, M., Dunham, M. J., Rettie, A. E., & Fowler, D. M. (2019). Applying multiplex assays to understand variation in pharmacogenes. *Clinical Pharmacology and Therapeutics*, *106*(2), 290–294. https://doi.org/10.1002/cpt.1468

Correia, M. A., Sinclair, P. R., & De Matteis, F. (2011). Cytochrome p450 regulation: The interplay between its heme and apoprotein moieties in synthesis, assembly, repair and disposal. *Drug Metabolism Reviews*, *43*(1), 1–26. https://doi.org/10.3109/03602532.2010.515222

Courchesne, W. E. (2002). Characterization of a novel, broad-based fungicidal activity for the antiarrhythmic drug amiodarone. *The Journal of Pharmacology and Experimental Therapeutics*, *300*(1), 195–199. https://doi.org/10.1124/jpet.300.1.195

Cravatt, B. F., Wright, A. T., & Kozarich, J. W. (2008). Activity-based protein profiling: From enzyme chemistry to proteomic chemistry. *Annual Review of Biochemistry*, *77*, 383–414. https://doi.org/10.1146/annurev.biochem.75.101304.124125

Cubillos, F. A., Louis, E. J., & Liti, G. (2009). Generation of a large set of genetically tractable haploid and diploid Saccharomyces strains. *FEMS Yeast Research*, *9*(8), 1217–1225. https://doi.org/10.1111/j.1567-1364.2009.00583.x

Dai, D., Wang, Y., Wang, S., Geng, P., Hu, L., Hu, G., & Cai, J. (2013). In vitro functional characterization of 37 CYP2C9 allelic isoforms found in Chinese Han population. *Acta Pharmacologica Sinica*, *34*(11), 1449–1456. https://doi.org/10.1038/aps.2013.123

Danielson, P. B. (2002). The cytochrome P450 superfamily: Biochemistry, evolution and drug metabolism in humans. *Current Drug Metabolism*, *3*(6), 561–597. https://doi.org/10.2174/1389200023337054

Dickmann, L. J., Locuson, C. W., Jones, J. P., & Rettie, A. E. (2004). Differential roles of Arg97, Asp293, and Arg108 in enzyme stability and substrate specificity of CYP2C9. *Molecular Pharmacology*, *65*(4), 842–850. https://doi.org/10.1124/mol.65.4.842

Donato, M. T., & Gómez-Lechón, M. J. (2013). Fluorescence-based screening of cytochrome P450 activities in intact cells. *Methods in Molecular Biology (Clifton, N.J.)*, *987*, 135–148. https://doi.org/10.1007/978-1-62703-321-3_12

Donato, M. T., Jiménez, N., Castell, J. V., & Gómez-Lechón, M. J. (2004). Fluorescence-based assays for screening nine cytochrome P450 (P450) activities in intact cells expressing individual human P450 enzymes. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *32*(7), 699–706. https://doi.org/10.1124/dmd.32.7.699

Dong, J., Wang, G., Zhang, C., Tan, H., Sun, X., Wu, M., & Xiao, D. (2013). A two-step integration method for seamless gene deletion in baker's yeast. *Analytical Biochemistry*, *439*(1), 30–36. https://doi.org/10.1016/j.ab.2013.04.005

Drăgan, C.-A., Peters, F. T., Bour, P., Schwaninger, A. E., Schaan, S. M., Neunzig, I., Widjaja, M., Zapp, J., Kraemer, T., Maurer, H. H., & Bureik, M. (2011). Convenient gram-scale metabolite synthesis by engineered fission yeast strains expressing functional human P450 systems. *Applied Biochemistry and Biotechnology*, *163*(8), 965–980. https://doi.org/10.1007/s12010-010-9100-3

Duconge, J., Cadilla, C. L., Windemuth, A., Kocherla, M., Gorowski, K., Seip, R. L., Bogaard, K., Renta, J. Y., Piovanetti, P., D'Agostino, D., Santiago-Borrero, P. J., & Ruaño, G. (2009). Prevalence of combinatorial cyp2c9 and vkorc1 genotypes in puerto ricans: Implications for warfarin management in hispanics. *Ethnicity & Disease*, *19*(4), 390–395.

Dunnenberger, H. M., Crews, K. R., Hoffman, J. M., Caudle, K. E., Broeckel, U., Howard, S. C., Hunkler, R. J., Klein, T. E., Evans, W. E., & Relling, M. V. (2015). Preemptive clinical pharmacogenetics implementation: Current programs in five united states medical centers. *Annual Review of Pharmacology and Toxicology*, *55*, 89–106. https://doi.org/10.1146/annurev-pharmtox-010814-124835

Emmerstorfer, A., Wimmer-Teubenbacher, M., Wriessnegger, T., Leitner, E., Müller, M., Kaluzna, I., Schürmann, M., Mink, D., Zellnig, G., Schwab, H., & Pichler, H. (2015). Over-expression of ICE2 stabilizes cytochrome P450 reductase in Saccharomyces cerevisiae and Pichia pastoris. *Biotechnology Journal*, *10*(4), 623–635. https://doi.org/10.1002/biot.201400780

Esposito, D., Weile, J., Shendure, J., Starita, L. M., Papenfuss, A. T., Roth, F. P., Fowler, D. M., & Rubin, A. F. (2019). MaveDB: An open-source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome Biology*, *20*(1), 223. https://doi.org/10.1186/s13059-019-1845-6

Evans, W. E., & Relling, M. V. (1999). Pharmacogenomics: Translating functional genomics into rational therapeutics. *Science (New York, N.Y.)*, *286*(5439), 487–491. https://doi.org/10.1126/science.286.5439.487

Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., Janizek, J. D., Huang, X., Starita, L. M., & Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. *Nature*, *562*(7726), 217–222. https://doi.org/10.1038/s41586-018-0461-z

Flora, D. R., Rettie, A. E., Brundage, R. C., & Tracy, T. S. (2017). Cyp2c9 genotype-dependent warfarin pharmacokinetics: Impact of cyp2c9 genotype on r- and s-warfarin and their oxidative metabolites. *Journal of Clinical Pharmacology*, *57*(3), 382–393. https://doi.org/10.1002/jcph.813

Foti, R. S., Wienkers, L. C., & Wahlstrom, J. L. (2010). Application of cytochrome P450 drug interaction screening in drug discovery. *Combinatorial Chemistry & High Throughput Screening*, *13*(2), 145–158. https://doi.org/10.2174/138620710790596718

Fowler, D. M., & Fields, S. (2014). Deep mutational scanning: A new style of protein science. *Nature Methods*, *11*(8), 801–807. https://doi.org/10.1038/nmeth.3027

Fujikura, K., Ingelman-Sundberg, M., & Lauschke, V. M. (2015). Genetic variation in the human cytochrome P450 supergene family. *Pharmacogenetics and Genomics*, *25*(12), 584–594. https://doi.org/10.1097/FPC.0000000000000172

Gaedigk, A., Ingelman-Sundberg, M., Miller, N. A., Leeder, J. S., Whirl-Carrillo, M., Klein, T. E., & PharmVar Steering Committee. (2018). The pharmacogene variation (Pharmvar) consortium:

Incorporation of the human cytochrome p450 (Cyp) allele nomenclature database. *Clinical Pharmacology and Therapeutics*, *103*(3), 399–401. https://doi.org/10.1002/cpt.910

Gaisne, M., Bécam, A. M., Verdière, J., & Herbert, C. J. (1999). A "natural" mutation in Saccharomyces cerevisiae strains derived from S288c affects the complex regulatory gene HAP1 (Cyp1). *Current Genetics*, *36*(4), 195–200. https://doi.org/10.1007/s002940050490

Gasperini, M., Starita, L., & Shendure, J. (2016). The power of multiplexed functional analysis of genetic variants. *Nature Protocols*, *11*(10), 1782–1787. https://doi.org/10.1038/nprot.2016.135

Gay, S. C., Roberts, A. G., & Halpert, J. R. (2010). Structural features of cytochromes p450 and ligands that affect drug metabolism as revealed by x-ray crystallography and nmr. *Future Medicinal Chemistry*, *2*(9), 1451–1468.

Gelman, H., Dines, J. N., Berg, J., Berger, A. H., Brnich, S., Hisama, F. M., James, R. G., Rubin, A. F., Shendure, J., Shirts, B., Fowler, D. M., Starita, L. M., & Brotman Baty Institute Mutational Scanning Working Group. (2019). Recommendations for the collection and use of multiplexed functional data for clinical variant interpretation. *Genome Medicine*, *11*(1), 85. https://doi.org/10.1186/s13073-019-0698-7

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

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. https://doi.org/10.1038/nprot.2007.13

Gonzalez, F. J., Kimura, S., Tamura, S., & Gelboin, H. V. (1991). Expression of mammalian cytochrome P450 using baculovirus. *Methods in Enzymology*, *206*, 93–99. https://doi.org/10.1016/0076-6879(91)06080-m

Gordon, A. S., Tabor, H. K., Johnson, A. D., Snively, B. M., Assimes, T. L., Auer, P. L., Ioannidis, J. P. A., Peters, U., Robinson, J. G., Sucheston, L. E., Wang, D., Sotoodehnia, N., Rotter, J. I., Psaty, B. M., Jackson, R. D., Herrington, D. M., O'Donnell, C. J., Reiner, A. P., Rich, S. S., … NHLBI GO Exome Sequencing Project. (2014). Quantifying rare, deleterious variation in 12 human cytochrome P450 drug-metabolism genes in a large-scale exome dataset. *Human Molecular Genetics*, *23*(8), 1957–1963. https://doi.org/10.1093/hmg/ddt588

Gottesman, O., Scott, S. A., Ellis, S. B., Overby, C. L., Ludtke, A., Hulot, J.-S., Hall, J., Chatani, K., Myers, K., Kannry, J. L., & Bottinger, E. P. (2013). The clipmerge pgx program: Clinical implementation of personalized medicine through electronic health records and genomics - pharmacogenomics. *Clinical Pharmacology and Therapeutics*, *94*(2), 214–217. https://doi.org/10.1038/clpt.2013.72

Goulding, R., Dawes, D., Price, M., Wilkie, S., & Dawes, M. (2015). Genotype-guided drug prescribing: A systematic review and meta-analysis of randomized control trials. British Journal of Clinical Pharmacology, 80(4), 868–877. https://doi.org/10.1111/bcp.12475

Grabowska, D., Jablonska-Skwiecinska, E., Plochocka, D., Chelstowska, A., Lewandowska, I., Witos, I., Majewska, Z., Rokicka-Milewska, R., & Burzynska, B. (2004). A novel mutation in the glucose-6-phosphate dehydrogenase gene in a subject with chronic nonspherocytic hemolytic anemia—Characterization of enzyme using yeast expression system and molecular modeling. *Blood Cells, Molecules & Diseases*, *32*(1), 124–130. https://doi.org/10.1016/j.bcmd.2003.11.001

Grimm, J. B., Heckman, L. M., & Lavis, L. D. (2013). The chemistry of small-molecule fluorogenic probes. *Progress in Molecular Biology and Translational Science*, *113*, 1–34. https://doi.org/10.1016/B978-0-12-386932-6.00001-6

Gröer, C., Busch, D., Patrzyk, M., Beyer, K., Busemann, A., Heidecke, C. D., Drozdzik, M., Siegmund, W., & Oswald, S. (2014). Absolute protein quantification of clinically relevant cytochrome P450

enzymes and UDP-glucuronosyltransferases by mass spectrometry-based targeted proteomics. Journal of Pharmaceutical and Biomedical Analysis, 100, 393–401. https://doi.org/10.1016/j.jpba.2014.08.016

Güldener, U., Heck, S., Fielder, T., Beinhauer, J., & Hegemann, J. H. (1996). A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Research*, *24*(13), 2519–2524. https://doi.org/10.1093/nar/24.13.2519

Hamann, T., & Møller, B. L. (2007). Improved cloning and expression of cytochrome P450s and cytochrome P450 reductase in yeast. *Protein Expression and Purification*, *56*(1), 121–127. https://doi.org/10.1016/j.pep.2007.06.007

Hanson, K. L., VandenBrink, B. M., Babu, K. N., Allen, K. E., Nelson, W. L., & Kunze, K. L. (2010). Sequential metabolism of secondary alkyl amines to metabolic-intermediate complexes: Opposing roles for the secondary hydroxylamine and primary amine metabolites of desipramine, (S)-fluoxetine, and N-desmethyldiltiazem. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *38*(6), 963–972. https://doi.org/10.1124/dmd.110.032391

Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A., & Deisenhofer, J. (1995). Structure and function of cytochromes P450: A comparative analysis of three crystal structures. *Structure (London, England: 1993)*, *3*(1), 41–62. https://doi.org/10.1016/s0969-2126(01)00134-4

He, M., Korzekwa, K. R., Jones, J. P., Rettie, A. E., & Trager, W. F. (1999). Structural forms of phenprocoumon and warfarin that are metabolized at the active site of CYP2C9. *Archives of Biochemistry and Biophysics*, *372*(1), 16–28. https://doi.org/10.1006/abbi.1999.1468

Hickman, M. J., & Winston, F. (2007). Heme levels switch the function of hap1 of saccharomyces cerevisiae between transcriptional activator and transcriptional repressor. *Molecular and Cellular Biology*, *27*(21), 7414–7424. https://doi.org/10.1128/MCB.00887-07

Higashi, M. K., Veenstra, D. L., Kondo, L. M., Wittkowsky, A. K., Srinouanprachanh, S. L., Farin, F. M., & Rettie, A. E. (2002). Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. JAMA, 287(13), 1690–1698. https://doi.org/10.1001/jama.287.13.1690

Hiratsuka, M. (2012). In vitro assessment of the allelic variants of cytochrome P450. *Drug Metabolism and Pharmacokinetics*, *27*(1), 68–84. https://doi.org/10.2133/dmpk.dmpk-11-rv-090

Hong, V., Steinmetz, N. F., Manchester, M., & Finn, M. G. (2010). Labeling live cells by copper-catalyzed alkyne-azide click chemistry. *Bioconjugate Chemistry*, *21*(10), 1912–1916. https://doi.org/10.1021/bc100272z

Hope, E. A., & Dunham, M. J. (2014). Ploidy-regulated variation in biofilm-related phenotypes in natural isolates of Saccharomyces cerevisiae. *G3 (Bethesda, Md.)*, *4*(9), 1773–1786. https://doi.org/10.1534/g3.114.013250

Huang, M., Bai, Y., Sjostrom, S. L., Hallström, B. M., Liu, Z., Petranovic, D., Uhlén, M., Joensson, H. N., Andersson-Svahn, H., & Nielsen, J. (2015). Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(34), E4689-4696. https://doi.org/10.1073/pnas.1506460112

Huang, M., Joensson, H. N., & Nielsen, J. (2018). High-throughput microfluidics for the screening of yeast libraries. *Methods in Molecular Biology (Clifton, N.J.)*, *1671*, 307–317. https://doi.org/10.1007/978-1-4939-7295-1_19

Ionescu, C., & Caira, M. R. (Eds.). (2005). Pathways of biotransformation—Phase i reactions. In Drug Metabolism: Current Concepts (pp. 41–128). Springer Netherlands. https://doi.org/10.1007/1-4020-4142-X_2

Iwamura, A., Fukami, T., Hosomi, H., Nakajima, M., & Yokoi, T. (2011). CYP2C9-mediated metabolic activation of losartan detected by a highly sensitive cell-based screening assay. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *39*(5), 838–846. https://doi.org/10.1124/dmd.110.037259

Jain, P. C., & Varadarajan, R. (2014). A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. *Analytical Biochemistry*, *449*, 90–98. https://doi.org/10.1016/j.ab.2013.12.002

Jean, P., Lopez-Garcia, P., Dansette, P., Mansuy, D., & Goldstein, J. L. (1996). Oxidation of tienilic acid by human yeast-expressed cytochromes P-450 2C8, 2C9, 2C18 and 2C19. Evidence that this drug is a mechanism-based inhibitor specific for cytochrome P-450 2C9. *European Journal of Biochemistry*, *241*(3), 797–804. https://doi.org/10.1111/j.1432-1033.1996.00797.x

Ji, Y., Skierka, J. M., Blommel, J. H., Moore, B. E., VanCuyk, D. L., Bruflat, J. K., Peterson, L. M., Veldhuizen, T. L., Fadra, N., Peterson, S. E., Lagerstedt, S. A., Train, L. J., Baudhuin, L. M., Klee, E. W., Ferber, M. J., Bielinski, S. J., Caraballo, P. J., Weinshilboum, R. M., & Black, J. L. (2016). Preemptive pharmacogenomic testing for precision medicine: A comprehensive analysis of five actionable pharmacogenomic genes using next-generation dna sequencing and a customized cyp2d6 genotyping cascade. *The Journal of Molecular Diagnostics: JMD*, *18*(3), 438–445. https://doi.org/10.1016/j.jmoldx.2016.01.003

Johnson, J. A., Caudle, K. E., Gong, L., Whirl-Carrillo, M., Stein, C. M., Scott, S. A., Lee, M. T., Gage, B. F., Kimmel, S. E., Perera, M. A., Anderson, J. L., Pirmohamed, M., Klein, T. E., Limdi, N. A., Cavallari, L. H., & Wadelius, M. (2017). Clinical pharmacogenetics implementation consortium (Cpic) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. Clinical Pharmacology and Therapeutics, 102(3), 397–404. https://doi.org/10.1002/cpt.668

Jones, E. W. (1991). Tackling the protease problem in Saccharomyces cerevisiae. *Methods in Enzymology*, *194*, 428–453. https://doi.org/10.1016/0076-6879(91)94034-a

Kachroo, A. H., Laurent, J. M., Yellman, C. M., Meyer, A. G., Wilke, C. O., & Marcotte, E. M. (2015). Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science (New York, N.Y.)*, *348*(6237), 921–925. https://doi.org/10.1126/science.aaa0769

Käppeli, O. (1986). Cytochromes P-450 of yeasts. *Microbiological Reviews*, *50*(3), 244–258.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., … MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7

Karst, F., & Lacroute, F. (1977). Ertosterol biosynthesis in Saccharomyces cerevisiae: Mutants deficient in the early steps of the pathway. *Molecular & General Genetics: MGG*, *154*(3), 269–277. https://doi.org/10.1007/BF00571282

Kazui, M., Nishiya, Y., Ishizuka, T., Hagihara, K., Farid, N. A., Okazaki, O., Ikeda, T., & Kurihara, A. (2010). Identification of the human cytochrome P450 enzymes involved in the two oxidative steps in the bioactivation of clopidogrel to its pharmacologically active metabolite. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *38*(1), 92–99. https://doi.org/10.1124/dmd.109.029132

Kemper, B. (2004). Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. *Toxicology and Applied Pharmacology*, *199*(3), 305–315. https://doi.org/10.1016/j.taap.2003.11.030

Kent, U. M., Juschyshyn, M. I., & Hollenberg, P. F. (2001). Mechanism-based inactivators as probes of cytochrome P450 structure and function. *Current Drug Metabolism*, *2*(3), 215–243. https://doi.org/10.2174/1389200013338478

Kim, H.-K., Lee, E. J., Lee, Y.-J., Kim, J., Kim, Y., Kim, K., Lee, S.-W., Chang, S., Lee, Y. J., Lee, J. W., Lee, W., Chun, S., Son, B. H., Jung, K. H., Kim, Y.-M., Min, W.-K., & Ahn, S.-H. (2020). Impact of proactive high-throughput functional assay data on BRCA1 variant interpretation in 3684 patients with breast or ovarian cancer. *Journal of Human Genetics*, *65*(3), 209–220. https://doi.org/10.1038/s10038-019-0713-2

Kimmel, S. E., French, B., Kasner, S. E., Johnson, J. A., Anderson, J. L., Gage, B. F., Rosenberg, Y. D., Eby, C. S., Madigan, R. A., McBane, R. B., Abdel-Rahman, S. Z., Stevens, S. M., Yale, S., Mohler, E. R., Fang, M. C., Shah, V., Horenstein, R. B., Limdi, N. A., Muldowney, J. A. S., … COAG Investigators. (2013). A pharmacogenetic versus a clinical algorithm for warfarin dosing. *The New England Journal of Medicine*, *369*(24), 2283–2293. https://doi.org/10.1056/NEJMoa1310669

Kirchheiner, J., Meineke, I., Steinbach, N., Meisel, C., Roots, I., & Brockmöller, J. (2003). Pharmacokinetics of diclofenac and inhibition of cyclooxygenases 1 and 2: No relationship to the CYP2C9 genetic polymorphism in humans. *British Journal of Clinical Pharmacology*, *55*(1), 51–61. https://doi.org/10.1046/j.1365-2125.2003.01712.x

Ko, J. W., Desta, Z., Soukhova, N. V., Tracy, T., & Flockhart, D. A. (2000). In vitro inhibition of the cytochrome P450 (Cyp450) system by the antiplatelet drug ticlopidine: Potent effect on CYP2C19 and CYP2D6. *British Journal of Clinical Pharmacology*, *49*(4), 343–351. https://doi.org/10.1046/j.1365-2125.2000.00175.x

Koren, G., Cairns, J., Chitayat, D., Gaedigk, A., & Leeder, S. J. (2006). Pharmacogenetics of morphine poisoning in a breastfed neonate of a codeine-prescribed mother. *Lancet (London, England)*, *368*(9536), 704. https://doi.org/10.1016/S0140-6736(06)69255-6

Kumar, V., Wahlstrom, J. L., Rock, D. A., Warren, C. J., Gorman, L. A., & Tracy, T. S. (2006). CYP2C9 inhibition: Impact of probe selection and pharmacogenetics on in vitro inhibition profiles. Drug Metabolism and Disposition: The Biological Fate of Chemicals, 34(12), 1966–1975. https://doi.org/10.1124/dmd.106.010926

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*(Database issue), D980–D985. https://doi.org/10.1093/nar/gkt1113

Lanza, A. M., Curran, K. A., Rey, L. G., & Alper, H. S. (2014). A condition-specific codon optimization approach for improved heterologous gene expression in Saccharomyces cerevisiae. *BMC Systems Biology*, *8*, 33. https://doi.org/10.1186/1752-0509-8-33

Laurent, J. M., Young, J. H., Kachroo, A. H., & Marcotte, E. M. (2016). Efforts to make and apply humanized yeast. Briefings in Functional Genomics, 15(2), 155–163. https://doi.org/10.1093/bfgp/elv041

Lazarou, J., Pomeranz, B. H., & Corey, P. N. (1998). Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA*, *279*(15), 1200–1205. https://doi.org/10.1001/jama.279.15.1200

Lee, C. R., Goldstein, J. A., & Pieper, J. A. (2002). Cytochrome P450 2C9 polymorphisms: A comprehensive review of the in-vitro and human data. *Pharmacogenetics*, *12*(3), 251–263. https://doi.org/10.1097/00008571-200204000-00010

Lee, C. R., Pieper, J. A., Frye, R. F., Hinderliter, A. L., Blaisdell, J. A., & Goldstein, J. A. (2003). Tolbutamide, flurbiprofen, and losartan as probes of CYP2C9 activity in humans. *Journal of Clinical Pharmacology*, *43*(1), 84–91. https://doi.org/10.1177/0091270002239710

Li, L., & Zhang, Z. (2016). Development and applications of the copper-catalyzed azide-alkyne cycloaddition (Cuaac) as a bioorthogonal reaction. *Molecules*, *21*(10). https://doi.org/10.3390/molecules21101393

Liachko, I., Bhaskar, A., Lee, C., Chung, S. C. C., Tye, B.-K., & Keich, U. (2010). A comprehensive genome-wide map of autonomously replicating sequences in a naive genome. *PLoS Genetics*, *6*(5). https://doi.org/10.1371/journal.pgen.1000946

Liti, G. (2015). The fascinating and secret wild life of the budding yeast S. cerevisiae. *ELife*, *4*. https://doi.org/10.7554/eLife.05835

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., Tsai, I. J., Bergman, C. M., Bensasson, D., O'Kelly, M. J. T., van Oudenaarden, A., Barton, D. B. H., Bailes, E., Nguyen Ba, A. N., Jones, M., … Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*, *458*(7236), 337–341. https://doi.org/10.1038/nature07743

Liu, H., Styles, C. A., & Fink, G. R. (1996). Saccharomyces cerevisiae s288c has a mutation in flo8, a gene required for filamentous growth. *Genetics*, *144*(3), 967–978.

López-Garcia, M. P., Dansette, P. M., & Mansuy, D. (1994). Thiophene derivatives as new mechanism-based inhibitors of cytochromes P-450: Inactivation of yeast-expressed human liver cytochrome P-450 2C9 by tienilic acid. *Biochemistry*, *33*(1), 166–175. https://doi.org/10.1021/bi00167a022

Louvel, H., Gillet-Markowska, A., Liti, G., & Fischer, G. (2014). A set of genetically diverged Saccharomyces cerevisiae strains with markerless deletions of multiple auxotrophic genes. *Yeast (Chichester, England)*, *31*(3), 91–101. https://doi.org/10.1002/yea.2991

Luzum, J. A., Pakyz, R. E., Elsey, A. R., Haidar, C. E., Peterson, J. F., Whirl-Carrillo, M., Handelman, S. K., Palmer, K., Pulley, J. M., Beller, M., Schildcrout, J. S., Field, J. R., Weitzel, K. W.,

Cooper-DeHoff, R. M., Cavallari, L. H., O'Donnell, P. H., Altman, R. B., Pereira, N., Ratain, M. J., … Pharmacogenomics Research Network Translational Pharmacogenetics Program. (2017). The pharmacogenomics research network translational pharmacogenetics program: Outcomes and metrics of pharmacogenetic implementations across diverse healthcare systems. *Clinical Pharmacology and Therapeutics*, *102*(3), 502–510. https://doi.org/10.1002/cpt.630

Maekawa, K., Adachi, M., Matsuzawa, Y., Zhang, Q., Kuroki, R., Saito, Y., & Shah, M. B. (2017). Structural basis of single-nucleotide polymorphisms in cytochrome p450 2c9. *Biochemistry*, *56*(41), 5476–5480. https://doi.org/10.1021/acs.biochem.7b00795

Martin, M. A., & Kroetz, D. L. (2013). Abacavir pharmacogenetics – from initial reports to standard of care. *Pharmacotherapy*, *33*(7), 765–775. https://doi.org/10.1002/phar.1278

Matreyek, K. A., Starita, L. M., Stephany, J. J., Martin, B., Chiasson, M. A., Gray, V. E., Kircher, M., Khechaduri, A., Dines, J. N., Hause, R. J., Bhatia, S., Evans, W. E., Relling, M. V., Yang, W., Shendure, J., & Fowler, D. M. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nature Genetics*, *50*(6), 874–882. https://doi.org/10.1038/s41588-018-0122-z

Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N., & Fowler, D. M. (2020). An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Research*, *48*(1), e1. https://doi.org/10.1093/nar/gkz910

Matsuzawa, T., Morita, T., Tanaka, N., Tohda, H., & Takegawa, K. (2011). Identification of a galactose-specific flocculin essential for non-sexual flocculation and filamentous growth in Schizosaccharomyces pombe. *Molecular Microbiology*, *82*(6), 1531–1544. https://doi.org/10.1111/j.1365-2958.2011.07908.x

McDonald, M. G., Au, N. T., & Rettie, A. E. (2015). P450-based drug-drug interactions of amiodarone and its metabolites: Diversity of inhibitory mechanisms. *Drug Metabolism and Disposition*, *43*(11), 1661–1669. https://doi.org/10.1124/dmd.115.065623

McDonald, M. G., Ray, S., Amorosi, C. J., Sitko, K. A., Kowalski, J. P., Paco, L., Nath, A., Gallis, B., Totah, R. A., Dunham, M. J., Fowler, D. M., & Rettie, A. E. (2017). Expression and functional characterization of breast cancer-associated cytochrome p450 4z1 in saccharomyces cerevisiae. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *45*(12), 1364–1371. https://doi.org/10.1124/dmd.117.078188

McDonald, M. G., & Rettie, A. E. (2007). Sequential metabolism and bioactivation of the hepatotoxin benzbromarone: Formation of glutathione adducts from a catechol intermediate. *Chemical Research in Toxicology*, *20*(12), 1833–1842. https://doi.org/10.1021/tx7001228

Mehta, A., Mason, P. J., & Vulliamy, T. J. (2000). Glucose-6-phosphate dehydrogenase deficiency. *Bailliere's Best Practice & Research. Clinical Haematology*, *13*(1), 21–38. https://doi.org/10.1053/beha.1999.0055

Melet, A., Assrir, N., Jean, P., Pilar Lopez-Garcia, M., Marques-Soares, C., Jaouen, M., Dansette, P. M., Sari, M. A., & Mansuy, D. (2003). Substrate selectivity of human cytochrome P450 2C9: Importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid. *Archives of Biochemistry and Biophysics*, *409*(1), 80–91. https://doi.org/10.1016/s0003-9861(02)00548-9

Merritt, J., Butz, J. A., Ogunnaike, B. A., & Edwards, J. S. (2005). Parallel analysis of mutant human glucose 6-phosphate dehydrogenase in yeast using PCR colonies. *Biotechnology and Bioengineering*, *92*(5), 519–531. https://doi.org/10.1002/bit.20726

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. https://doi.org/10.1073/pnas.1212287109

Miners, J. O., & Birkett, D. J. (1998). Cytochrome P4502C9: An enzyme of major importance in human drug metabolism. *British Journal of Clinical Pharmacology*, *45*(6), 525–538. https://doi.org/10.1046/j.1365-2125.1998.00721.x

Miners, J. O., Wing, L. M., & Birkett, D. J. (1985). Normal metabolism of debrisoquine and theophylline in a slow tolbutamide metaboliser. Australian and New Zealand Journal of Medicine, 15(3), 348–349. https://doi.org/10.1111/j.1445-5994.1985.tb04052.x

Minowa, O., Sogawa, K., Higashi, Y., & Fujii-Kuriyama, Y. (1990). Functional expression of microsomal and mitochondrial cytochrome P-450 (D and scc) in COS-7 cells from cloned cDNA. *Cell Structure and Function*, *15*(1), 21–30. https://doi.org/10.1247/csf.15.21

Mosher, C. M., Tai, G., & Rettie, A. E. (2009). Cyp2c9 amino acid residues influencing phenytoin turnover and metabolite regio- and stereochemistry. *Journal of Pharmacology and Experimental Therapeutics*, *329*(3), 938–944. https://doi.org/10.1124/jpet.109.150706

Motulsky, A. G. (1957). Drug reactions enzymes, and biochemical genetics. *Journal of the American Medical Association*, *165*(7), 835–837. https://doi.org/10.1001/jama.1957.72980250010016

Mumberg, D., Müller, R., & Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, *156*(1), 119–122. https://doi.org/10.1016/0378-1119(95)00037-7

Nelson, D. R., Zeldin, D. C., Hoffman, S. M. G., Maltais, L. J., Wain, H. M., & Nebert, D. W. (2004). Comparison of cytochrome P450 (Cyp) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics*, *14*(1), 1–18. https://doi.org/10.1097/00008571-200401000-00001

Neunzig, I., Widjaja, M., Peters, F. T., Maurer, H. H., Hehn, A., Bourgaud, F., & Bureik, M. (2013). Coexpression of CPR from various origins enhances biotransformation activity of human CYPs in S. pombe. *Applied Biochemistry and Biotechnology*, *170*(7), 1751–1766. https://doi.org/10.1007/s12010-013-0303-2

Niinuma, Y., Saito, T., Takahashi, M., Tsukada, C., Ito, M., Hirasawa, N., & Hiratsuka, M. (2014). Functional characterization of 32 CYP2C9 allelic variants. *The Pharmacogenomics Journal*, *14*(2), 107–114. https://doi.org/10.1038/tpj.2013.22

Novartis Pharmaceuticals Corporation. Mayzent (siponimod) [package insert]. U.S. Food and Drug Administration website. https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/209884s000lbl.pdf. Revised March 2019. Accessed August , 2020

O'Donnell, J. P., Dalvie, D. K., Kalgutkar, A. S., & Obach, R. S. (2003). Mechanism-based inactivation of human recombinant P450 2C9 by the nonsteroidal anti-inflammatory drug suprofen. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *31*(11), 1369–1377. https://doi.org/10.1124/dmd.31.11.1369

Oeda, K., Sakaki, T., & Ohkawa, H. (1985). Expression of rat liver cytochrome P-450MC cDNA in Saccharomyces cerevisiae. *DNA (Mary Ann Liebert, Inc.)*, *4*(3), 203–210. https://doi.org/10.1089/dna.1985.4.203

Owen, R. P., Gong, L., Sagreiya, H., Klein, T. E., & Altman, R. B. (2010). Vkorc1 pharmacogenomics summary. *Pharmacogenetics and Genomics*, *20*(10), 642–644. https://doi.org/10.1097/FPC.0b013e32833433b6

Pelkonen, P., Lang, M. A., Wild, C. P., Negishi, M., & Juvonen, R. O. (1994). Activation of aflatoxin B1 by mouse CYP2A enzymes and cytotoxicity in recombinant yeast cells. *European Journal of Pharmacology*, *292*(1), 67–73. https://doi.org/10.1016/0926-6917(94)90027-2

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., … Schacherer, J. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature*, *556*(7701), 339–344. https://doi.org/10.1038/s41586-018-0030-5

Peyronneau, M. A., Renaud, J. P., Truan, G., Urban, P., Pompon, D., & Mansuy, D. (1992). Optimization of yeast-expressed human liver cytochrome P450 3A4 catalytic activities by coexpressing NADPH-cytochrome P450 reductase and cytochrome b5. *European Journal of Biochemistry*, *207*(1), 109–116. https://doi.org/10.1111/j.1432-1033.1992.tb17027.x

Pirmohamed, M., Burnside, G., Eriksson, N., Jorgensen, A. L., Toh, C. H., Nicholson, T., Kesteven, P., Christersson, C., Wahlström, B., Stafberg, C., Zhang, J. E., Leathart, J. B., Kohnke, H., Maitland-van der Zee, A. H., Williamson, P. R., Daly, A. K., Avery, P., Kamali, F., Wadelius, M., & EU-PACT Group. (2013). A randomized trial of genotype-guided dosing of warfarin. *The New England Journal of Medicine*, *369*(24), 2294–2303. https://doi.org/10.1056/NEJMoa1311386

Pirmohamed, M., Kamali, F., Daly, A. K., & Wadelius, M. (2015). Oral anticoagulation: A critique of recent advances and controversies. *Trends in Pharmacological Sciences*, *36*(3), 153–163. https://doi.org/10.1016/j.tips.2015.01.003

Pompon, D., Louerat, B., Bronine, A., & Urban, P. (1996). Yeast expression of animal and plant P450s in optimized redox environments. *Methods in Enzymology*, *272*, 51–64. https://doi.org/10.1016/s0076-6879(96)72008-6

Pompon, D., Perret, A., Bellamine, A., Laine, R., Gautier, J. C., & Urban, P. (1995). Genetically engineered yeast cells and their applications. *Toxicology Letters*, *82–83*, 815–822. https://doi.org/10.1016/0378-4274(95)03522-2

Pompon, D., Truan, G., Bellamine, A., & Urban, P. (1994). Expression of cytochromes p450 in yeast: Practical aspects. In M. R. Waterman & M. Hildebrand (Eds.), *Assessment of the Use of Single Cytochrome P450 Enzymes in Drug Research* (pp. 97–110). Springer. https://doi.org/10.1007/978-3-662-03019-6_6

Porter, T. D. (2002). The roles of cytochrome b5 in cytochrome P450 reactions. *Journal of Biochemical and Molecular Toxicology*, *16*(6), 311–316. https://doi.org/10.1002/jbt.10052

Porubsky, P. R., Meneely, K. M., & Scott, E. E. (2008). Structures of human cytochrome P-450 2E1. Insights into the binding of inhibitors and both small molecular weight and fatty acid substrates. *The Journal of Biological Chemistry*, *283*(48), 33698–33707. https://doi.org/10.1074/jbc.M805999200

Pulley, J. M., Denny, J. C., Peterson, J. F., Bernard, G. R., Vnencak-Jones, C. L., Ramirez, A. H., Delaney, J. T., Bowton, E., Brothers, K., Johnson, K., Crawford, D. C., Schildcrout, J., Masys, D. R., Dilks, H. H., Wilke, R. A., Clayton, E. W., Shultz, E., Laposata, M., McPherson, J., … Roden, D. M. (2012). Operational implementation of prospective genotyping for personalized medicine: The design of the Vanderbilt PREDICT project. *Clinical Pharmacology and Therapeutics*, *92*(1), 87–95. https://doi.org/10.1038/clpt.2011.371

Rademacher, P. M., Woods, C. M., Huang, Q., Szklarz, G. D., & Nelson, S. D. (2012). Differential oxidation of two thiophene-containing regioisomers to reactive metabolites by cytochrome p450 2c9. *Chemical Research in Toxicology*, *25*(4), 895–903. https://doi.org/10.1021/tx200519d

Ratcliff, W. C., Denison, R. F., Borrello, M., & Travisano, M. (2012). Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1595–1600. https://doi.org/10.1073/pnas.1115323109

Relling, M. V., & Klein, T. E. (2011). Cpic: Clinical pharmacogenetics implementation consortium of the pharmacogenomics research network. *Clinical Pharmacology and Therapeutics*, *89*(3), 464–467. https://doi.org/10.1038/clpt.2010.279

Relling, Mary V., & Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature*, *526*(7573), 343–350. https://doi.org/10.1038/nature15817

Relling, M. V., Schwab, M., Whirl‐Carrillo, M., Suarez‐Kurtz, G., Pui, C.-H., Stein, C. M., Moyer, A. M., Evans, W. E., Klein, T. E., Antillon‐Klussmann, F. G., Caudle, K. E., Kato, M., Yeoh, A. E. J., Schmiegelow, K., & Yang, J. J. (2019). Clinical pharmacogenetics implementation consortium guideline for thiopurine dosing based on tpmt and nudt15 genotypes: 2018 update. Clinical Pharmacology & Therapeutics, 105(5), 1095–1105. https://doi.org/10.1002/cpt.1304

Rettie, A. E., Korzekwa, K. R., Kunze, K. L., Lawrence, R. F., Eddy, A. C., Aoyama, T., Gelboin, H. V., Gonzalez, F. J., & Trager, W. F. (1992). Hydroxylation of warfarin by human cDNA-expressed cytochrome P-450: A role for P-4502C9 in the etiology of (S)-warfarin-drug interactions. Chemical Research in Toxicology, 5(1), 54–59. https://doi.org/10.1021/tx00025a009

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., & ACMG Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Rieder, M. J., Reiner, A. P., Gage, B. F., Nickerson, D. A., Eby, C. S., McLeod, H. L., Blough, D. K., Thummel, K. E., Veenstra, D. L., & Rettie, A. E. (2005). Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. The New England Journal of Medicine, 352(22), 2285–2293. https://doi.org/10.1056/NEJMoa044503

Rieger, M. A., Ebner, R., Bell, D. R., Kiessling, A., Rohayem, J., Schmitz, M., Temme, A., Rieber, E. P., & Weigle, B. (2004). Identification of a novel mammary-restricted cytochrome P450, CYP4Z1, with overexpression in breast carcinoma. Cancer Research, 64(7), 2357–2364. https://doi.org/10.1158/0008-5472.can-03-0849

Roberts, E. S., Hopkins, N. E., Alworth, W. L., & Hollenberg, P. F. (1993). Mechanism-based inactivation of cytochrome P450 2B1 by 2-ethynylnaphthalene: Identification of an active-site peptide. *Chemical Research in Toxicology*, *6*(4), 470–479. https://doi.org/10.1021/tx00034a013

Rubin, A. F., Gelman, H., Lucas, N., Bajjalieh, S. M., Papenfuss, A. T., Speed, T. P., & Fowler, D. M. (2017). A statistical framework for analyzing deep mutational scanning data. *Genome Biology*, *18*(1), 150. https://doi.org/10.1186/s13059-017-1272-5

Sardi, M., & Gasch, A. P. (2017). Incorporating comparative genomics into the design–test–learn cycle of microbial strain engineering. *FEMS Yeast Research*, *17*(5). https://doi.org/10.1093/femsyr/fox042

Schmid, R. D., & Urlacher, V. (2007). *Modern biooxidation: Enzymes, reactions and applications*. John Wiley & Sons.

Scordo, M. G., Pengo, V., Spina, E., Dahl, M. L., Gusella, M., & Padrini, R. (2002). Influence of CYP2C9 and CYP2C19 genetic polymorphisms on warfarin maintenance dose and metabolic clearance. *Clinical Pharmacology and Therapeutics*, *72*(6), 702–710. https://doi.org/10.1067/mcp.2002.129321

Shepelin, D., Hansen, A. S. L., Lennen, R., Luo, H., & Herrgård, M. J. (2018). Selecting the best: Evolutionary engineering of chemical production in microbes. *Genes*, *9*(5). https://doi.org/10.3390/genes9050249

Shibata, M., Sakaki, T., Yabusaki, Y., Murakami, H., & Ohkawa, H. (1990). Genetically engineered P450 monooxygenases: Construction of bovine P450c17/yeast reductase fused enzymes. *DNA and Cell Biology*, *9*(1), 27–36. https://doi.org/10.1089/dna.1990.9.27

Sim, S. C., & Ingelman-Sundberg, M. (2010). The Human Cytochrome P450 (Cyp) Allele Nomenclature website: A peer-reviewed database of CYP variants and their associated effects. *Human Genomics*, *4*(4), 278–281. https://doi.org/10.1186/1479-7364-4-4-278

Sim, S. C., Risinger, C., Dahl, M.-L., Aklillu, E., Christensen, M., Bertilsson, L., & Ingelman-Sundberg, M. (2006). A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. *Clinical Pharmacology and Therapeutics*, *79*(1), 103–113. https://doi.org/10.1016/j.clpt.2005.10.002

Sridar, C., Goosen, T. C., Kent, U. M., Williams, J. A., & Hollenberg, P. F. (2004). Silybin inactivates cytochromes P450 3A4 and 2C9 and inhibits major hepatic glucuronosyltransferases. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *32*(6), 587–594. https://doi.org/10.1124/dmd.32.6.587

Stansfield, I., & Kelly, S. L. (1996). Purification and quantification of Saccharomyces cerevisiae cytochrome P450. *Methods in Molecular Biology (Clifton, N.J.)*, *53*, 355–366. https://doi.org/10.1385/0-89603-319-8:355

Starita, L. M., Ahituv, N., Dunham, M. J., Kitzman, J. O., Roth, F. P., Seelig, G., Shendure, J., & Fowler, D. M. (2017). Variant interpretation: Functional assays to the rescue. *American Journal of Human Genetics*, *101*(3), 315–325. https://doi.org/10.1016/j.ajhg.2017.07.014

Starita, L. M., Young, D. L., Islam, M., Kitzman, J. O., Gullingsrud, J., Hause, R. J., Fowler, D. M., Parvin, J. D., Shendure, J., & Fields, S. (2015). Massively parallel functional analysis of brca1 ring domain variants. *Genetics*, *200*(2), 413–422. https://doi.org/10.1534/genetics.115.175802

Stearns, V., Johnson, M. D., Rae, J. M., Morocho, A., Novielli, A., Bhargava, P., Hayes, D. F., Desta, Z., & Flockhart, D. A. (2003). Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *Journal of the National Cancer Institute*, *95*(23), 1758–1764. https://doi.org/10.1093/jnci/djg108

Steensels, J., Snoek, T., Meersman, E., Nicolino, M. P., Voordeckers, K., & Verstrepen, K. J. (2014). Improving industrial yeast strains: Exploiting natural and artificial diversity. *Fems Microbiology Reviews*, *38*(5), 947–995. https://doi.org/10.1111/1574-6976.12073

Steward, D. J., Haining, R. L., Henne, K. R., Davis, G., Rushmore, T. H., Trager, W. F., & Rettie, A. E. (1997). Genetic association between sensitivity to warfarin and expression of CYP2C9*3. Pharmacogenetics, 7(5), 361–367. https://doi.org/10.1097/00008571-199710000-00004

Stubbins, M. J., Harries, L. W., Smith, G., Tarbit, M. H., & Wolf, C. R. (1996). Genetic analysis of the human cytochrome P450 CYP2C9 locus. Pharmacogenetics, 6(5), 429–439. https://doi.org/10.1097/00008571-199610000-00007

Suiter, C. C., Moriyama, T., Matreyek, K. A., Yang, W., Scaletti, E. R., Nishii, R., Yang, W., Hoshitsuki, K., Singh, M., Trehan, A., Parish, C., Smith, C., Li, L., Bhojwani, D., Yuen, L. Y. P., Li, C.-K., Li, C.-H., Yang, Y.-L., Walker, G. J., … Yang, J. J. (2020). Massively parallel variant characterization identifies NUDT15 alleles associated with thiopurine toxicity. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(10), 5394–5401. https://doi.org/10.1073/pnas.1915680117

Sullivan-Klose, T. H., Ghanayem, B. I., Bell, D. A., Zhang, Z. Y., Kaminsky, L. S., Shenfield, G. M., Miners, J. O., Birkett, D. J., & Goldstein, J. A. (1996). The role of the CYP2C9-Leu359 allelic variant in the tolbutamide polymorphism. *Pharmacogenetics*, *6*(4), 341–349. https://doi.org/10.1097/00008571-199608000-00007

Sun, S., Weile, J., Verby, M., Wu, Y., Wang, Y., Cote, A. G., Fotiadou, I., Kitaygorodsky, J., Vidal, M., Rine, J., Ješina, P., Kožich, V., & Roth, F. P. (2020). A proactive genotype-to-patient-phenotype map for cystathionine beta-synthase. *Genome Medicine*, *12*(1), 13. https://doi.org/10.1186/s13073-020-0711-1

Sultana, J., Cutroneo, P., & Trifirò, G. (2013). Clinical and economic burden of adverse drug reactions. Journal of Pharmacology & Pharmacotherapeutics, 4(Suppl1), S73–S77. https://doi.org/10.4103/0976-500X.120957

Tai, G., Farin, F., Rieder, M. J., Dreisbach, A. W., Veenstra, D. L., Verlinde, C. L. M. J., & Rettie, A. E. (2005). In-vitro and in-vivo effects of the CYP2C9*11 polymorphism on warfarin metabolism and dose. *Pharmacogenetics and Genomics*, *15*(7), 475–481. https://doi.org/10.1097/01.fpc.0000162005.80857.98

Takanashi, K., Tainaka, H., Kobayashi, K., Yasumori, T., Hosakawa, M., & Chiba, K. (2000). CYP2C9 Ile359 and Leu359 variants: Enzyme kinetic study with seven substrates. *Pharmacogenetics*, *10*(2), 95–104. https://doi.org/10.1097/00008571-200003000-00001

Theberge, A. B., Courtois, F., Schaerli, Y., Fischlechner, M., Abell, C., Hollfelder, F., & Huck, W. T. S. (2010). Microdroplets in microfluidics: An evolving platform for discoveries in chemistry and biology. *Angewandte Chemie (International Ed. in English)*, *49*(34), 5846–5868. https://doi.org/10.1002/anie.200906653

Theken, K. N., Lee, C. R., Gong, L., Caudle, K. E., Formea, C. M., Gaedigk, A., Klein, T. E., Agúndez, J. A. G., & Grosser, T. (2020). Clinical pharmacogenetics implementation consortium guideline (Cpic) for cyp2c9 and nonsteroidal anti-inflammatory drugs. *Clinical Pharmacology and Therapeutics*, *108*(2), 191–200. https://doi.org/10.1002/cpt.1830

Thorn, C. F., Klein, T. E., & Altman, R. B. (2013). Pharmgkb: The pharmacogenomics knowledge base. *Methods in Molecular Biology (Clifton, N.J.)*, *1015*, 311–320. https://doi.org/10.1007/978-1-62703-435-7_20

Trubetskoy, O. V., Gibson, J. R., & Marks, B. D. (2005). Highly miniaturized formats for in vitro drug metabolism assays using vivid fluorescent substrates and recombinant human cytochrome P450 enzymes. *Journal of Biomolecular Screening*, *10*(1), 56–66. https://doi.org/10.1177/1087057104269731

Ung, Y. T., Ong, C. E., & Pan, Y. (2018). Current high-throughput approaches of screening modulatory effects of xenobiotics on cytochrome p450 (Cyp) enzymes. *High-Throughput*, *7*(4). https://doi.org/10.3390/ht7040029

Urban, P., Truan, G., Bellamine, A., Laine, R., Gautier, J. C., & Pompon, D. (1994). Engineered yeasts simulating P450-dependent metabolisms: Tricks, myths and reality. *Drug Metabolism and Drug Interactions*, *11*(3), 169–200. https://doi.org/10.1515/dmdi.1994.11.3.169

Urban, P., Truan, G., Gautier, J. C., & Pompon, D. (1993). Xenobiotic metabolism in humanized yeast: Engineered yeast cells producing human NADPH-cytochrome P-450 reductase, cytochrome b5, epoxide hydrolase and P-450s. *Biochemical Society Transactions*, *21*(4), 1028–1034. https://doi.org/10.1042/bst0211028

van der Hoeven, T. A. (1981). Isolation of hepatic microsomes by polyethylene glycol 6000 fractionation of the postmitochondrial fraction. *Analytical Biochemistry*, *115*(2), 398–402. https://doi.org/10.1016/0003-2697(81)90024-5

van Leeuwen, J. S., Vredenburg, G., Dragovic, S., Tjong, T. F. J., Vos, J. C., & Vermeulen, N. P. E. (2011). Metabolism related toxicity of diclofenac in yeast as model system. *Toxicology Letters*, *200*(3), 162–168. https://doi.org/10.1016/j.toxlet.2010.11.010

Veronese, M. E., Mackenzie, P. I., Doecke, C. J., McManus, M. E., Miners, J. O., & Birkett, D. J. (1991). Tolbutamide and phenytoin hydroxylations by cDNA-expressed human liver cytochrome P4502C9. Biochemical and Biophysical Research Communications, 175(3), 1112–1118. https://doi.org/10.1016/0006-291x(91)91680-b

Vogl, S., Lutz, R. W., Schönfelder, G., & Lutz, W. K. (2015). CYP2C9 genotype vs. Metabolic phenotype for individual drug dosing—A correlation analysis using flurbiprofen as probe drug. *PloS One*, *10*(3), e0120403. https://doi.org/10.1371/journal.pone.0120403

Wada, Y., Mitsuda, M., Ishihara, Y., Watanabe, M., Iwasaki, M., & Asahi, S. (2008). Important amino acid residues that confer CYP2C19 selective activity to CYP2C9. Journal of Biochemistry, 144(3), 323–333. https://doi.org/10.1093/jb/mvn070

Walsky, R. L., & Obach, R. S. (2004). Validated assays for human cytochrome P450 activities. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *32*(6), 647–660. https://doi.org/10.1124/dmd.32.6.647

Wang, H., An, N., Wang, H., Gao, Y., Liu, D., Bian, T., Zhu, J., & Chen, C. (2011). Evaluation of the effects of 20 nonsynonymous single nucleotide polymorphisms of CYP2C19 on S-mephenytoin 4'-hydroxylation and omeprazole 5'-hydroxylation. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *39*(5), 830–837. https://doi.org/10.1124/dmd.110.037549

Wang, Y.-H., Pan, P.-P., Dai, D.-P., Wang, S.-H., Geng, P.-W., Cai, J.-P., & Hu, G.-X. (2014). Effect of 36 CYP2C9 variants found in the Chinese population on losartan metabolism in vitro. *Xenobiotica; the Fate of Foreign Compounds in Biological Systems*, *44*(3), 270–275. https://doi.org/10.3109/00498254.2013.820007

Warringer, J., Zörgö, E., Cubillos, F. A., Zia, A., Gjuvsland, A., Simpson, J. T., Forsmark, A., Durbin, R., Omholt, S. W., Louis, E. J., Liti, G., Moses, A., & Blomberg, A. (2011). Trait variation in yeast

is defined by population history. *PLoS Genetics*, *7*(6), e1002111. https://doi.org/10.1371/journal.pgen.1002111

Waterman, M. R. (1993). Heterologous expression of cytochrome P-450 in Escherichia coli. *Biochemical Society Transactions*, *21*(4), 1081–1085. https://doi.org/10.1042/bst0211081

Weile, J., & Roth, F. P. (2018). Multiplexed assays of variant effects contribute to a growing genotype-phenotype atlas. *Human Genetics*, *137*(9), 665–678. https://doi.org/10.1007/s00439-018-1916-x

Weile, J., Sun, S., Cote, A. G., Knapp, J., Verby, M., Mellor, J. C., Wu, Y., Pons, C., Wong, C., van Lieshout, N., Yang, F., Tasan, M., Tan, G., Yang, S., Fowler, D. M., Nussbaum, R., Bloom, J. D., Vidal, M., Hill, D. E., … Roth, F. P. (2017). A framework for exhaustively mapping functional missense variants. *Molecular Systems Biology*, *13*(12). https://doi.org/10.15252/msb.20177908

Wennerholm, A., Johansson, I., Hidestrand, M., Bertilsson, L., Gustafsson, L. L., & Ingelman-Sundberg, M. (2001). Characterization of the CYP2D6*29 allele commonly present in a black Tanzanian population causing reduced catalytic activity. *Pharmacogenetics*, *11*(5), 417–427. https://doi.org/10.1097/00008571-200107000-00005

Wester, M. R., Yano, J. K., Schoch, G. A., Yang, C., Griffin, K. J., Stout, C. D., & Johnson, E. F. (2004). The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. *The Journal of Biological Chemistry*, *279*(34), 35630–35637. https://doi.org/10.1074/jbc.M405427200

Wiek, C., Schmidt, E. M., Roellecke, K., Freund, M., Nakano, M., Kelly, E. J., Kaisers, W., Yarov-Yarovoy, V., Kramm, C. M., Rettie, A. E., & Hanenberg, H. (2015). Identification of amino acid determinants in CYP4B1 for optimal catalytic processing of 4-ipomeanol. *The Biochemical Journal*, *465*(1), 103–114. https://doi.org/10.1042/BJ20140813

Wilkening, S., Lin, G., Fritsch, E. S., Tekkedil, M. M., Anders, S., Kuehn, R., Nguyen, M., Aiyar, R. S., Proctor, M., Sakhanenko, N. A., Galas, D. J., Gagneur, J., Deutschbauer, A., & Steinmetz, L. M.

(2014). An evaluation of high-throughput approaches to QTL mapping in Saccharomyces cerevisiae. *Genetics*, *196*(3), 853–865. https://doi.org/10.1534/genetics.113.160291

Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Matak Vinković, D., & Jhoti, H. (2003). Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature*, *424*(6947), 464–468. https://doi.org/10.1038/nature01862

Winkler, J. D., & Kao, K. C. (2014). Recent advances in the evolutionary engineering of industrial biocatalysts. *Genomics*, *104*(6 Pt A), 406–411. https://doi.org/10.1016/j.ygeno.2014.09.006

Winterstein, A. G., Sauer, B. C., Hepler, C. D., & Poole, C. (2002). Preventable drug-related hospital admissions. The Annals of Pharmacotherapy, 36(7–8), 1238–1248. https://doi.org/10.1345/aph.1A225

Woronoff, G., El Harrak, A., Mayot, E., Schicke, O., Miller, O. J., Soumillion, P., Griffiths, A. D., & Ryckelynck, M. (2011). New generation of amino coumarin methyl sulfonate-based fluorogenic substrates for amidase assays in droplet-based microfluidic applications. *Analytical Chemistry*, *83*(8), 2852–2857. https://doi.org/10.1021/ac200373n

Wright, A. T., & Cravatt, B. F. (2007). Chemical proteomic probes for profiling cytochrome p450 activities and drug interactions in vivo. *Chemistry & Biology*, *14*(9), 1043–1051. https://doi.org/10.1016/j.chembiol.2007.08.008

Wright, A. T., Song, J. D., & Cravatt, B. F. (2009). A suite of activity-based probes for human cytochrome P450 enzymes. *Journal of the American Chemical Society*, *131*(30), 10692–10700. https://doi.org/10.1021/ja9037609

Yabusaki, Y. (1995). Artificial P450/reductase fusion enzymes: What can we learn from their structures? *Biochimie*, *77*(7–8), 594–603. https://doi.org/10.1016/0300-9084(96)88175-2

Yamazaki, H., Nakamura, M., Komatsu, T., Ohyama, K., Hatanaka, N., Asahi, S., Shimada, N., Guengerich, F. P., Shimada, T., Nakajima, M., & Yokoi, T. (2002). Roles of NADPH-P450

reductase and apo- and holo-cytochrome b5 on xenobiotic oxidations catalyzed by 12 recombinant human cytochrome P450s expressed in membranes of Escherichia coli. *Protein Expression and Purification*, *24*(3), 329–337. https://doi.org/10.1006/prep.2001.1578

Yang, Y., Botton, M. R., Scott, E. R., & Scott, S. A. (2017). Sequencing the CYP2D6 gene: From variant allele discovery to clinical pharmacogenetic testing. *Pharmacogenomics*, *18*(7), 673–685. https://doi.org/10.2217/pgs-2017-0033

Yuan, H.-Y., Chen, J.-J., Lee, M. T. M., Wung, J.-C., Chen, Y.-F., Charng, M.-J., Lu, M.-J., Hung, C.-R., Wei, C.-Y., Chen, C.-H., Wu, J.-Y., & Chen, Y.-T. (2005). A novel functional VKORC1 promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity. *Human Molecular Genetics*, *14*(13), 1745–1751. https://doi.org/10.1093/hmg/ddi180

Zanger, U. M., & Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics*, *138*(1), 103–141. https://doi.org/10.1016/j.pharmthera.2012.12.007

Zhang, H.-F., Wang, H.-H., Gao, N., Wei, J.-Y., Tian, X., Zhao, Y., Fang, Y., Zhou, J., Wen, Q., Gao, J., Zhang, Y.-J., Qian, X.-H., & Qiao, H.-L. (2016). Physiological content and intrinsic activities of 10 cytochrome p450 isoforms in human normal liver microsomes. *The Journal of Pharmacology and Experimental Therapeutics*, *358*(1), 83–93. https://doi.org/10.1124/jpet.116.233635

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics (Oxford, England)*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

Zhang, L., Sarangi, V., Moon, I., Yu, J., Liu, D., Devarajan, S., Reid, J. M., Kalari, K. R., Wang, L., & Weinshilboum, R. (2020). Cyp2c9 and cyp2c19: Deep mutational scanning and functional characterization of genomic missense variants. *Clinical and Translational Science*, *13*(4), 727–742. https://doi.org/10.1111/cts.12758

Zhou, D., Lu, Y., Steiner, M. S., & Dalton, J. T. (2000). Cytochrome P-450 2C9 sensitizes human prostate tumor cells to cyclophosphamide via a bystander effect. *Antimicrobial Agents and Chemotherapy*, *44*(10), 2659–2663. https://doi.org/10.1128/aac.44.10.2659-2663.2000

Zhou, Y., Ingelman‐Sundberg, M., & Lauschke, V. (2017). Worldwide distribution of cytochrome p450 alleles: A meta‐analysis of population‐scale sequencing projects. *Clinical Pharmacology and Therapeutics*, *102*(4), 688–700. https://doi.org/10.1002/cpt.690

Appendix

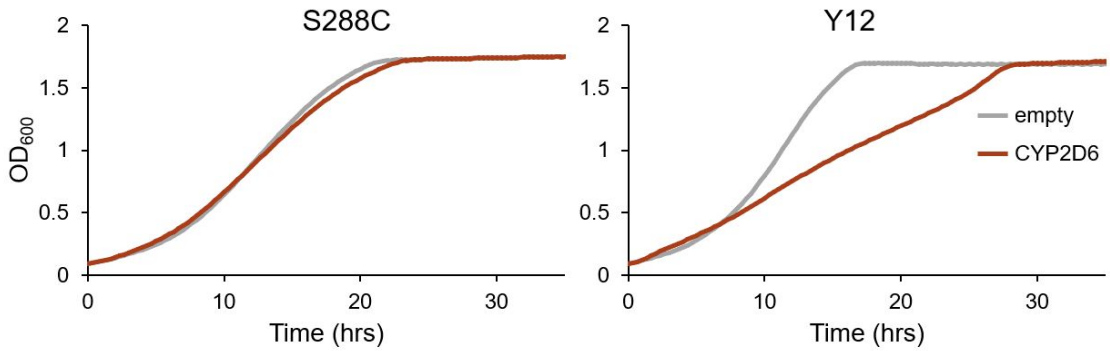

Figure A.1. CYP2D6-dependent growth inhibition of Y12 strain. Growth profiles of S288C strain YMD4353 (left) and Y12 strain (right) expressing empty vector in grey or CYP2D6 in red. Cells growth in galactose media at 30 °C.

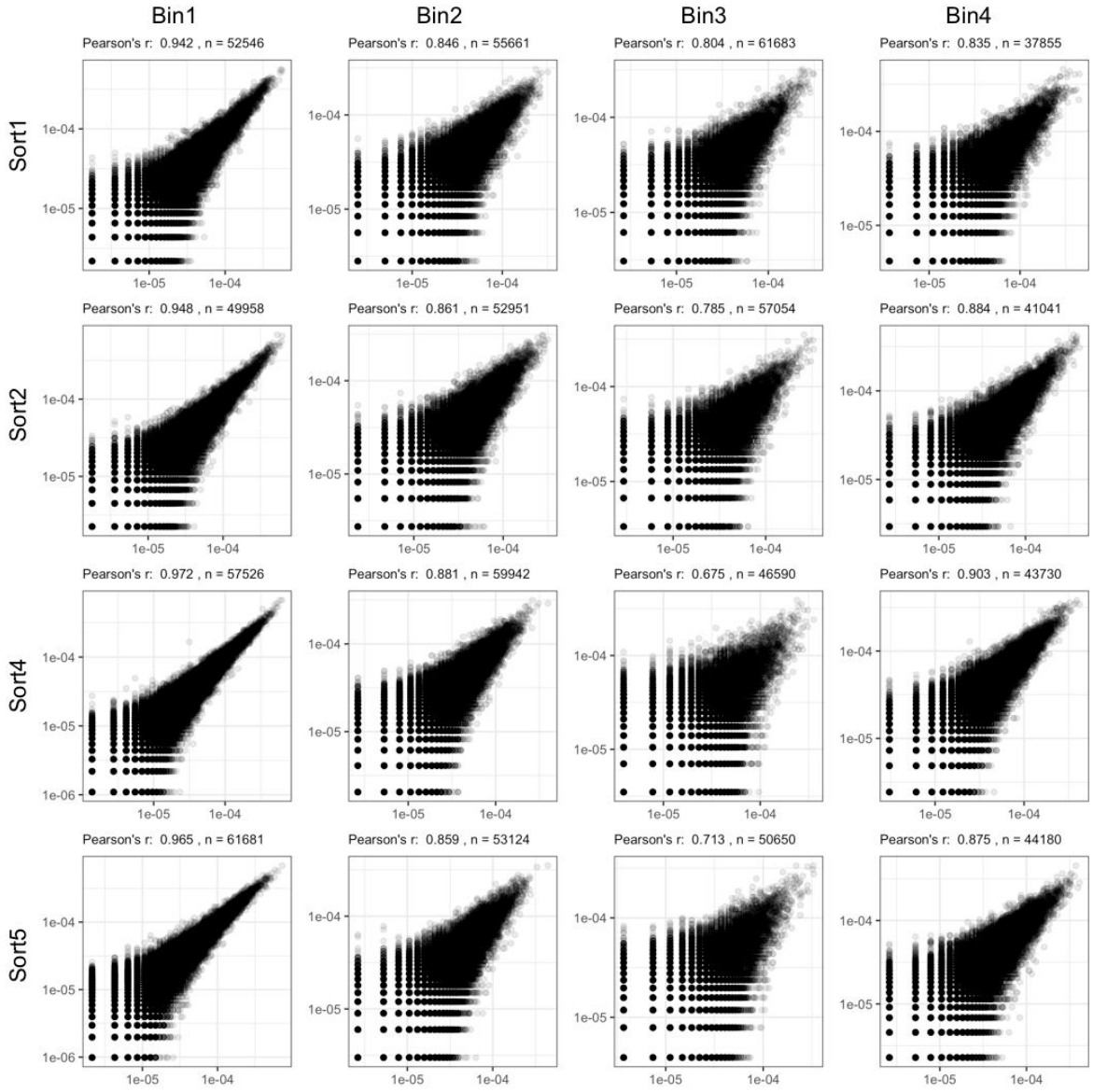

Figure A.2. CYP2C9 activity library technical replicate correlation. Sequencing of technical (PCR) replicates of CYP2C9 activity library: Scatterplots of barcode frequency correlation of for each bin for each of the four sorts of the CYP2C9 activity library

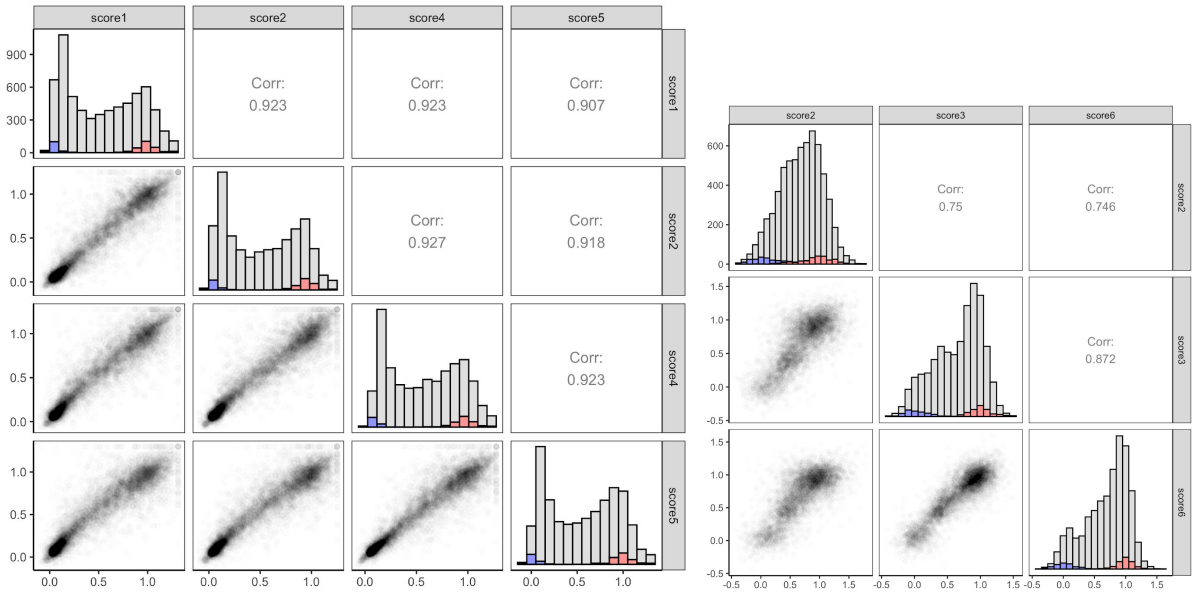

Figure A.3. CYP2C9 score correlation matrices. Replicate correlation of CYP2C9 activity scores for the four replicates, and CYP2C9 abundance scores for the three replicates. Bottom corner: pairwise scatterplot of scores, diagonal: stacked histograms of synonymous (red), missense (grey), and nonsense (blue) score distributions, top corner: Pearson's r values.

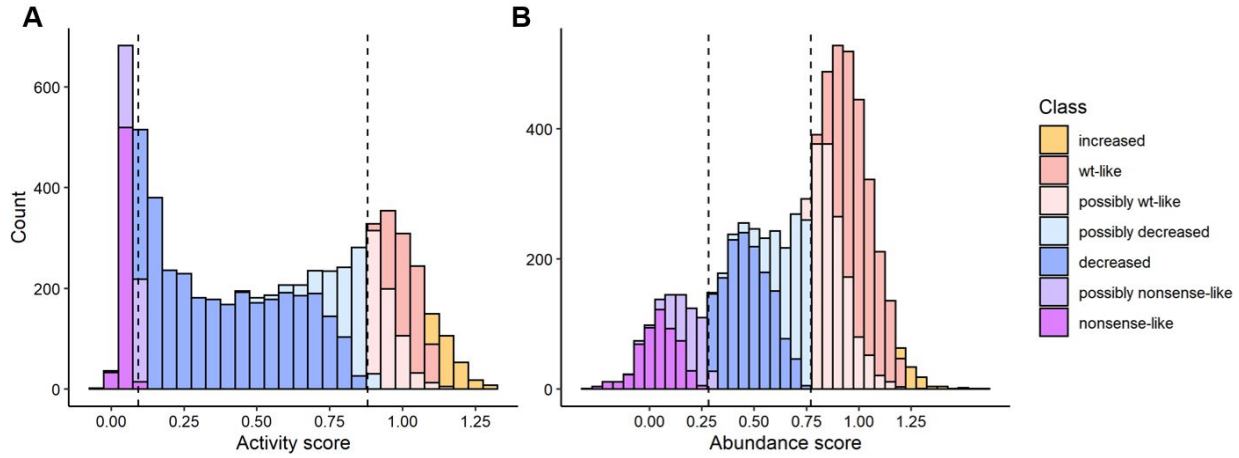

Figure A.4 Classification of CYP2C9 scores into classes. Stacked histograms of CYP2C9 activity and abundance scores categorized into functional classes. In dotted lines, the 95th percentile of the nonsense distribution (left), and the 5th percentile of the synonymous distribution (right), used for categorization. Variants were categorized by determining whether variant scores and confidence intervals fell within the synonymous and nonsense variant thresholds, as detailed in Methods

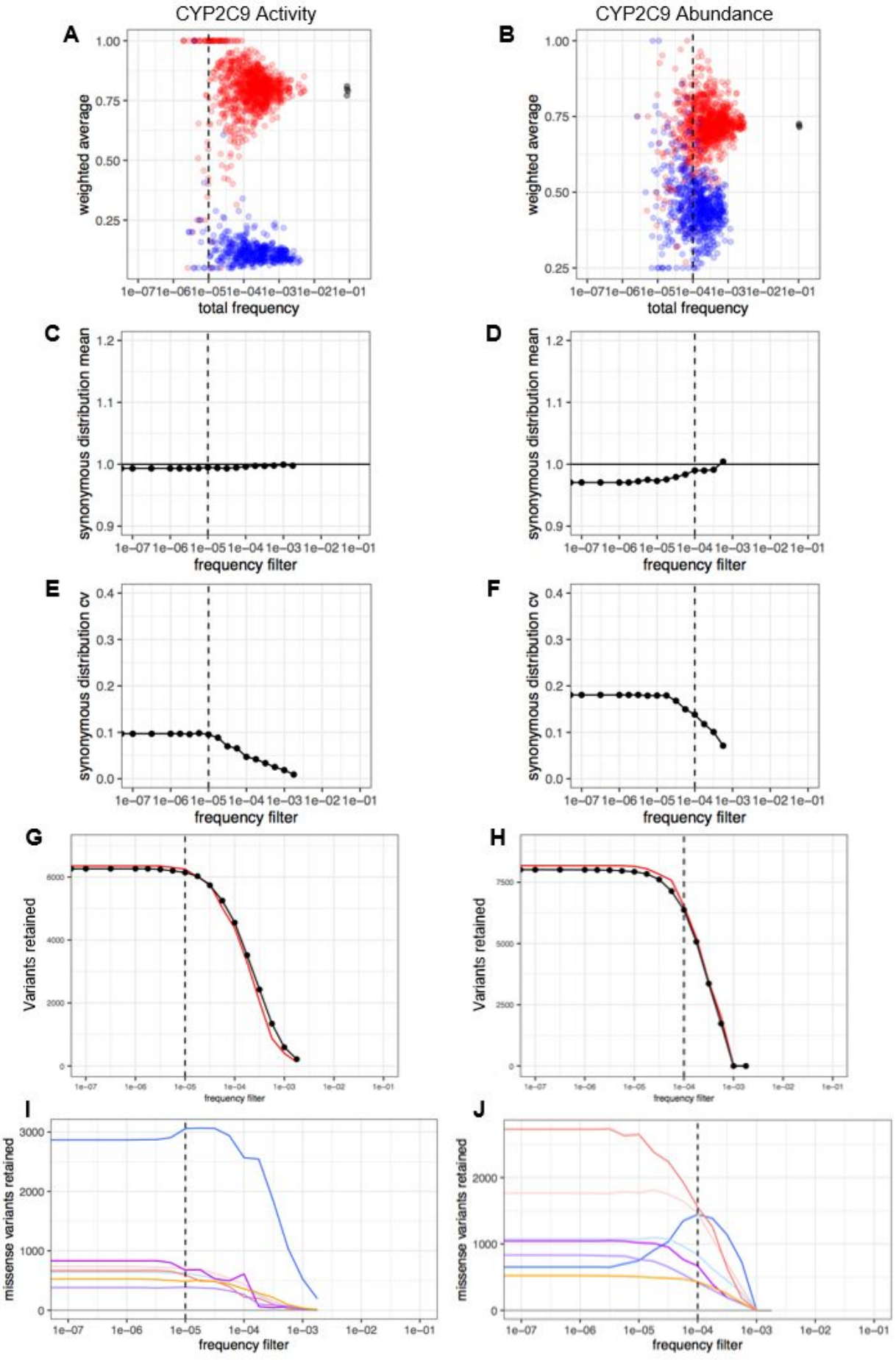
CYP2C9 Activity
CYP2C9 Abundance
A
B
C
D
E
F
G
H
I
J
weighted average
total frequency
synonymous distribution mean
synonymous distribution cv
frequency filter
Variants retained
missense variants retained
1e−07
1e−06
1e−05
1e−04
1e−03
1e−02
1e−01

Figure A.5. Determining variant frequency filters. In a and b), scatterplots of variant weighted average vs total frequency for wildtype (black), synonymous (red), and nonsense (blue) variants, for each of the four or three replicates for the CYP2C9 activity and abundance libraries, respectively. In c and d), the mean of the synonymous distribution at different frequency filters for the activity and abundance library, respectively. In e and f), coefficient of variation of the synonymous distribution at varying frequency filters. In g and h), the number of missense variants retained (black) at varying frequency filters, and 25 times the number of synonymous variants retained shown in red. In i and j), the number of missense variants retained in each functional class at varying frequency filters. Classes are nonsense-like (purple), possibly nonsense-like (lilac), decreased (blue), possibly decreased (light blue), possibly wt-like (salmon), wt-like (red), increased (orange), and unknown (grey). For all plots, the frequency filter used for library analysis is shown as a dashed line, for the activity library, this is 10e-5, and for the abundance library, this is 10e-4.

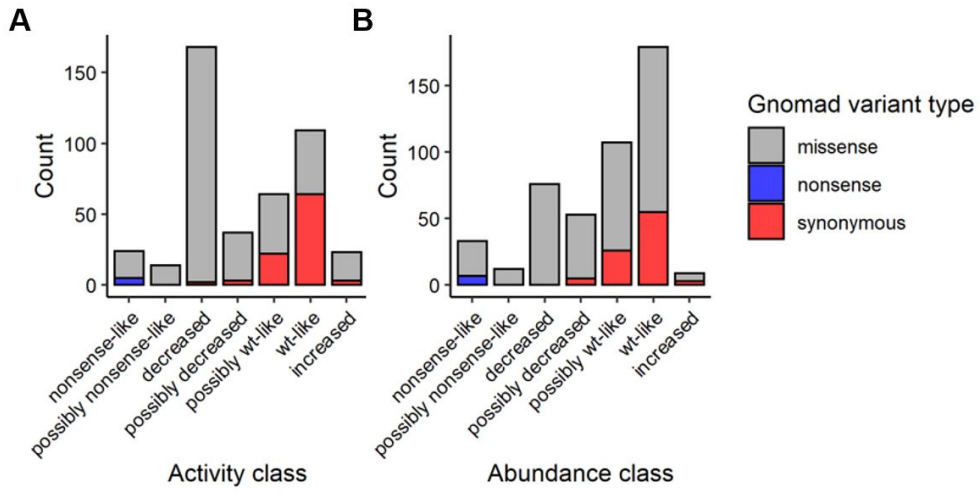

Figure A.6. GnomAD distribution in activity and abundance classes. Bar plot of variant activity and abundance class, colored by type of mutation in GnomAD. GnomAD variants combined from v2 and v3 data, and filtered for missense, stop-gained, and missense variants. A total of 439 and 469 GnomAD variants are shown with activity and abundance classes, respectively.

Table A.1. CYP2C9 library fluorescence-activated cell sorts.
Four-way sorts of the yeast activity CYP2C9 library and the HEK 293T human abundance CYP2C9 library. For the yeast activity library, the approximate binning target percentages were: Bin1: 60%, Bin2: 20%, Bin3: 10%, Bin4: 10%. The human abundance library was binned into equal 25% bins. Unless otherwise noted in the experiment number column, DNA amplified directly from sorted cells, rather than growing out culture and then amplifying. Asterisk indicates approximate cell numbers.

Library	Experiment number	Cells sorted in Bin1	Cells sorted in Bin2	Cells sorted in Bin3	Cells sorted in Bin4
Yeast activity	1	16,519,559	8,327,656	4,167,071	3,868,841
Yeast activity	2	13,747,214	5,500,208	3,056,205	3,165,603
Yeast activity	4	18,155,749	5,665,686	2,901,456	3,279,558
Yeast activity	5	15,405,689	5,377,714	3,040,153	3,005,108
Human abundance	2	625,888	573,840	624,090	542,210
Human abundance	3 (growout)	1,000,000*	1,000,000*	1,000,000*	1,000,000*
Human abundance	6 (growout)	1,000,000*	1,000,000*	1,000,000*	1,000,000*

Table A.2. Library subassembly statistics.
Barcoded CYP2C9 libraries sequenced on a Sequel II with chemistry TODO. CCS reads (circular consensus reads) generated with ccs2. During subassembly, barcodes with only two ccs reads were aligned and where sequences differed, the base from the read with the higher quality score was used. Human abundance library sequenced from two replicate sampleings of the same E. coli ligation transformation.

Library	**Yeast activity**	**Human abundance**
SMRT cells	2	2
CCS reads with 10 or more passes	309,948	545,277
CCS reads passing filters (mapping, soft clipping, correct length barcode)	283,195	515,420

Unique barcodes	105,372 (2.9x)	78,740 (6.9x)
Barcodes with one consensus read	41,218	7,376
Barcodes with two consensus reads	23,806	8,252
Barcodes with three or more consensus reads	40,348	63,112
Barcodes with identical consensus reads	51,348	28,578
Barcodes assigned with multiple sequence alignment	Major allele: 16,906 Quality: 37,118	Major allele: 42,805 Quality: 7,357
Barcodes associated with WT CYP2C9 sequence (or synonymous mutation)	2,974	3,697
Barcodes associated with single amino acid mutation	38,127	49,015
# single amino acid mutations	**6,542 (66.8%)**	**8,310 (84.8%)**
Barcodes associated with indel	54,385	18,436
Barcodes associated with two or more amino acid mutation	9,886	7,592
# unique nucleotide sequences	66,958	37,758
# unique full length nucleotide sequences	22,421	22,669

Table A.3. CYP2C9 Star Allele CPIC functional annotations

Star allele: star number assigned to each CYP2C9 allele by PharmVar, Variant: amino acid change, CPIC functional status: allele functional status determined by the Clinical Pharmacogenetics Implementation Consortium (CPIC), CPIC evidence: evidence level for allele functional status (definitive, stong, moderate, limited, or inadequate evidence), Activity score: variant score determined by Click-seq, Activity sd: activity score standard deviation, Activity class: functional classification determined from activity score.

Star	**Variant**	**CPIC**	**CPIC**	**Activity**	**Activity**	**Activity class**

allele		functional status	evidence	score	sd	
*1	WT	normal function	NA	0.985	0.015	wt-like
*2	R144C	decreased function	definitive	0.834	0.294	possibly_decreased
*3	I359L	decreased function	definitive	0.445	0.039	decreased
*4	I359T	decreased function	NA	0.928	0.281	possibly_wt-like
*5	D360E	decreased function	NA	NA	NA	NA
*7	L19I	uncertain function	NA	NA	NA	NA
*8	R150H	decreased function	definitive	0.732	0.112	decreased
*9	H251R	normal function	strong	NA	NA	NA
*10	E272G	uncertain function	inadequate evidence	NA	NA	NA
*11	R335W	decreased function	definitive	0.108	0.024	decreased
*12	P489S	decreased function	NA	0.153	0.024	decreased
*13	L90P	decreased function	definitive	0.413	0.266	decreased
*14	R125H	decreased function	moderate	0.648	0.063	decreased
*15	S162X	no function	limited	0.039	0.04	nonsense-like
*16	T299A	decreased function	limited	0.624	0.167	decreased

*17	P382S	uncertain function	inadequate evidence	0.308	0.054	decreased
*19	Q454H	uncertain function	inadequate evidence	0.257	0.025	decreased
*20	G70R	uncertain function	inadequate evidence	NA	NA	NA
*21	P30L	uncertain function	inadequate evidence	0.1	0.018	decreased
*22	N41D	uncertain function	inadequate evidence	NA	NA	NA
*23	V76M	decreased function	limited	0.202	0.07	decreased
*24	E354K	no function	limited	0.047	0.041	nonsense-like
*26	T130R	decreased function	limited	NA	NA	NA
*27	R150L	uncertain function	inadequate evidence	0.713	0.071	decreased
*28	Q214L	decreased function	limited	0.844	0.024	decreased
*29	P279T	decreased function	limited	NA	NA	NA
*30	A477T	decreased function	limited	0.909	0.091	possibly_wt-like
*31	I327T	decreased function	limited	0.264	0.062	decreased
*32	V490F	uncertain function	inadequate evidence	0.608	0.021	decreased
*33	R132Q	decreased function	limited	NA	NA	NA
*34	R335Q	uncertain function	inadequate evidence	0.6	0.076	decreased

*36	M1V	uncertain function	inadequate evidence	NA	NA	NA
*37	D49G	decreased function	limited	0.922	0.067	possibly_wt-like
*38	G96A	decreased function	limited	0.933	0.06	possibly_wt-like
*39	G98V	no function	limited	NA	NA	NA
*40	F110S	uncertain function	inadequate evidence	0.781	0.079	decreased
*41	K119R	uncertain function	inadequate evidence	1.045	0.047	wt-like
*42	R124Q	no function	limited	0.517	0.021	decreased
*43	R124W	no function	limited	0.154	0.077	decreased
*44	T130M	decreased function	limited	NA	NA	NA
*45	R132W	no function	limited	0.375	0.073	decreased
*46	A149T	decreased function	limited	0.105	0.007	decreased
*47	P163L	uncertain function	inadequate evidence	0.293	0.038	decreased
*48	I207T	uncertain function	inadequate evidence	0.413	0.102	decreased
*49	I222V	uncertain function	inadequate evidence	NA	NA	NA
*50	P227S	decreased function	limited	0.759	0.206	possibly_decreased
*51	I284V	uncertain function	inadequate evidence	0.775	0.075	decreased
*52	T299R	no function	limited	0.481	0.037	decreased

*53	P317S	uncertain function	inadequate evidence	0.5	0.43	possibly_decreased
*54	S343R	uncertain function	inadequate evidence	1.002	0.106	wt-like
*55	L361I	decreased function	limited	0.567	0.123	decreased
*56	I387V	uncertain function	inadequate evidence	1.016	0.039	wt-like
*57	N204H	unknown function	NA	0.359	0.087	decreased
*58	P337T	decreased function	inadequate evidence	0.451	0.035	decreased
*59	I434F	decreased function	inadequate evidence	1.111	0.062	increased
*60	L467P	decreased function	inadequate evidence	0.045	0.028	nonsense-like

VITA

Clara Amorosi was born in Seattle, Washington. She attended Harvey Mudd College and received a Bachelor of Sciences degree in Mathematical and Computational Biology. During her undergraduate studies, she conducted research under Dr. Daniel Stoebel on how the sigma factor RpoS affects *E. coli* metabolism. She was a member of the Amgen Scholars summer internship program in 2013 and conducted research on microbiome scaling laws under Dr. Elhanan Borenstein at the University of Washington Department of Genome Sciences. She joined the Genome Sciences doctoral program in the fall of 2014 and conducted her thesis work with Dr. Maitreya Dunham, developing yeast-based assays for studying human P450 activity in a massively parallel format. She has presented her work at numerous scientific conferences and been involved in departmental outreach activities. She served as a teaching assistant at the Cold Spring Harbor Laboratory Yeast Genetics and Genomics course in 2017.

Publications

Hope, E. A., **Amorosi, C. J.**, Miller, A. W., Dang, K., Heil, C. S., & Dunham, M. J. (2017). Experimental evolution reveals favored adaptive routes to cell aggregation in yeast. *Genetics*, *206*(2), 1153–1167. https://doi.org/10.1534/genetics.116.198895

McDonald, M. G., Ray, S., **Amorosi, C. J.**, Sitko, K. A., Kowalski, J. P., Paco, L., Nath, A., Gallis, B., Totah, R. A., Dunham, M. J., Fowler, D. M., & Rettie, A. E. (2017). Expression and functional characterization of breast cancer-associated cytochrome p450 4z1 in saccharomyces cerevisiae. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *45*(12), 1364–1371. https://doi.org/10.1124/dmd.117.078188