The American Journal of Human Genetics, Volume *108*

Supplemental information

Massively parallel characterization of CYP2C9 variant enzyme activity and abundance

Clara J. Amorosi, Melissa A. Chiasson, Matthew G. McDonald, Lai Hong Wong, Katherine A. Sitko, Gabriel Boyle, John P. Kowalski, Allan E. Rettie, Douglas M. Fowler, and Maitreya J. Dunham

SUPPLEMENTAL FIGURES

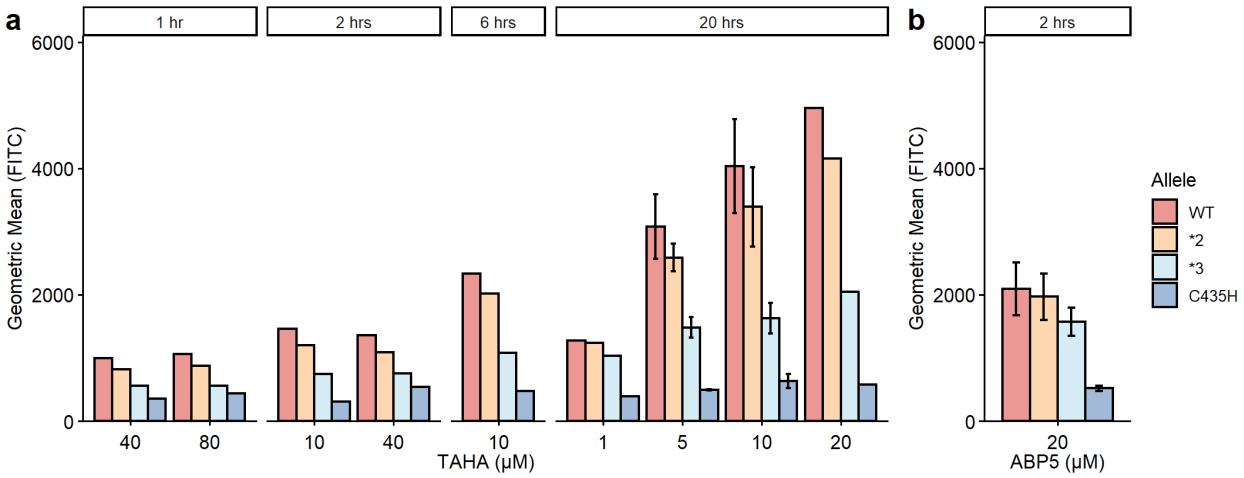


Figure S1. Probe labeling optimization of CYP2C9 alleles.
Barplot of flow cytometry of ABPP-labeled CYP2C9 WT (red), reduced activity alleles (*2 and *3, orange and turquoise), and null allele (C435H, blue). Cells labeled with TAHA probe (a) or ABP5 probe (b). Incubation times tested shown on top, probe concentration shown on bottom (μM). Error bars show standard deviation, each sample from ~20,000 cells.

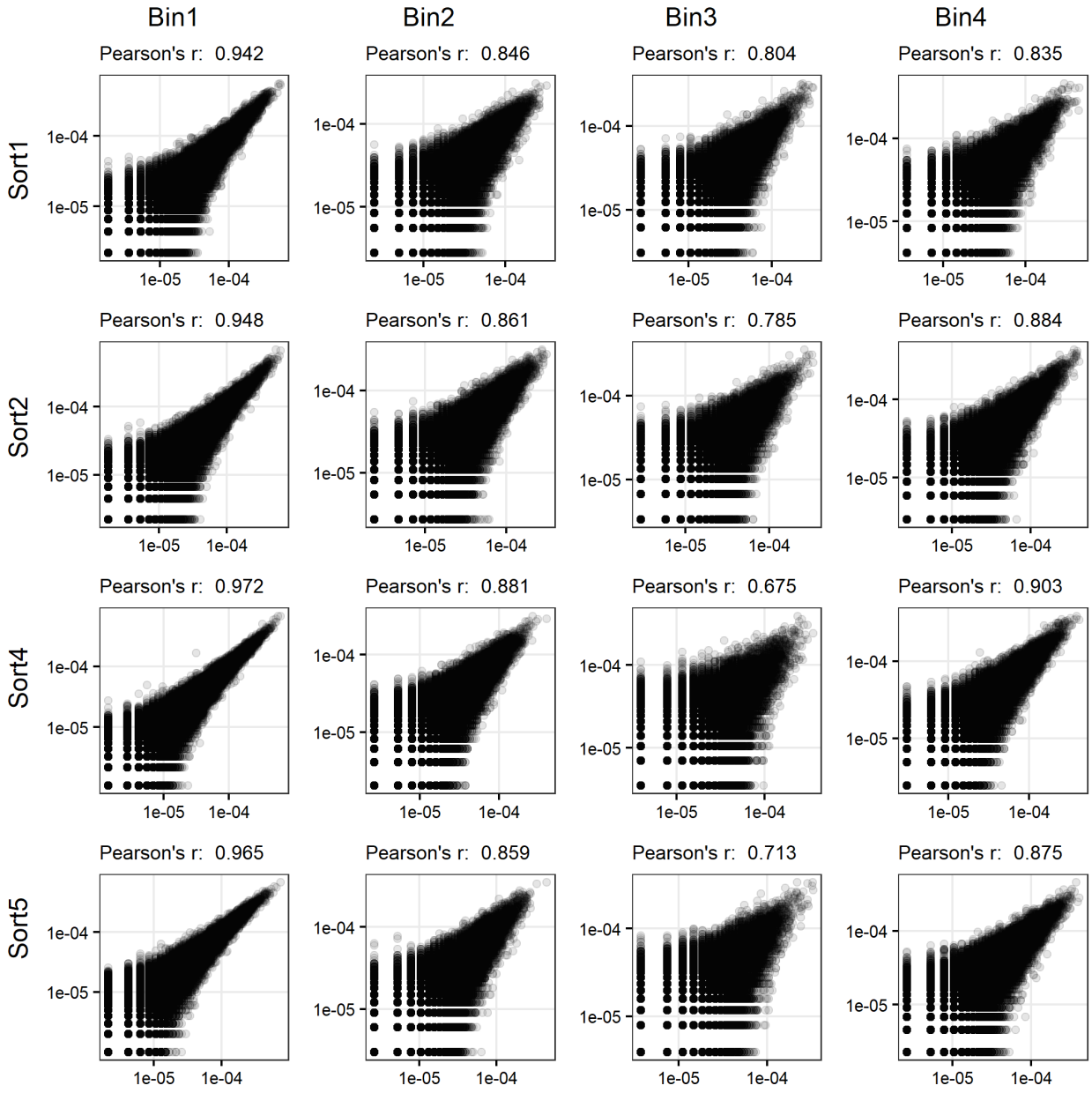


Figure S2. CYP2C9 activity library technical replicate correlation.
Sequencing of technical (PCR) replicates of CYP2C9 activity library: Scatterplots of barcode frequency correlation of each bin for each of the four sorts of the CYP2C9 activity library.

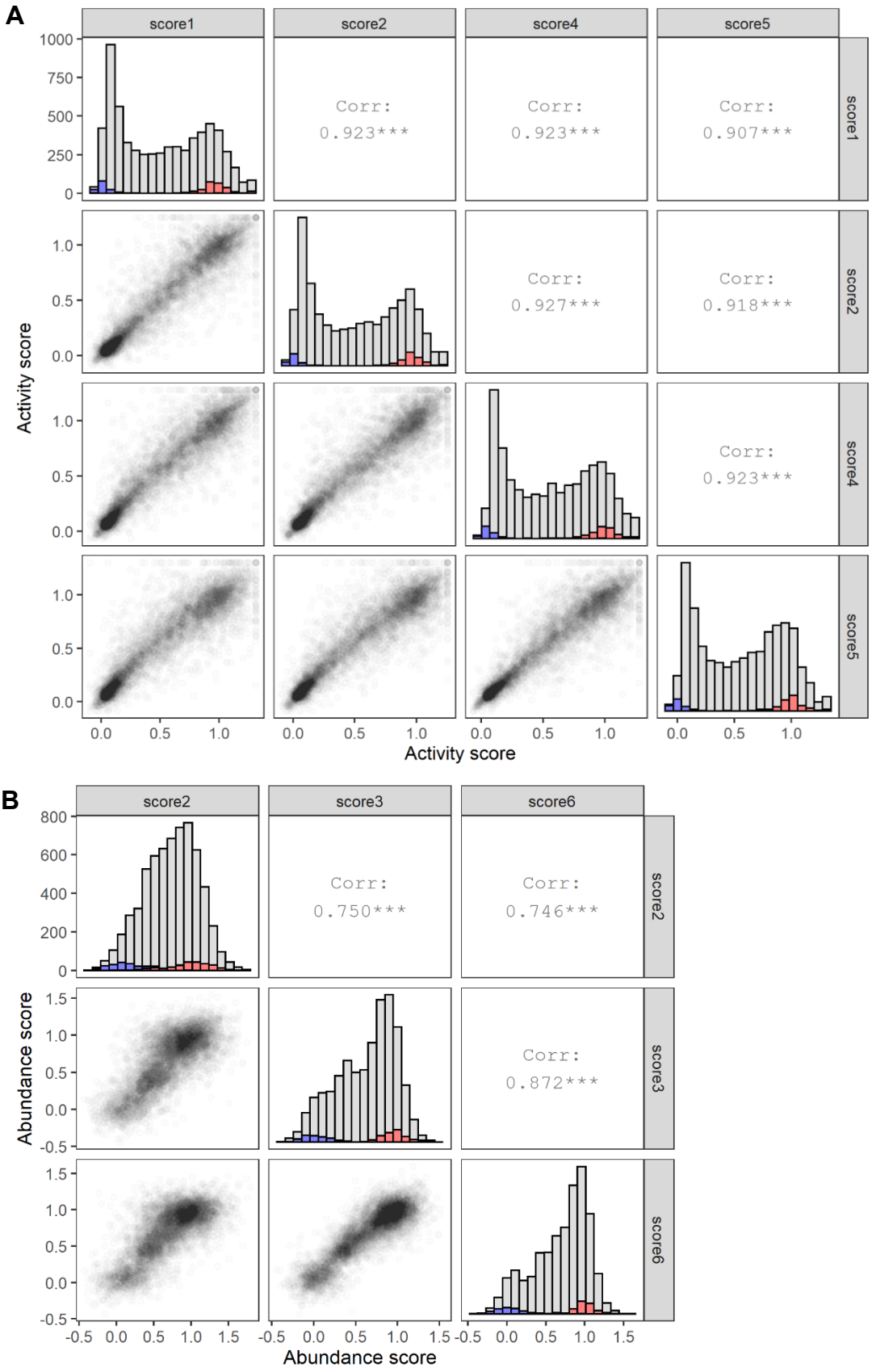


Figure S3. CYP2C9 score correlation matrices.
Replicate correlation of CYP2C9 activity scores for the four replicates (top), and CYP2C9 abundance scores for the three replicates (bottom). Bottom corner: pairwise

scatterplot of scores, diagonal: stacked histograms of synonymous (red), missense (grey), and nonsense (blue) score distributions, top corner: Pearson's r values.

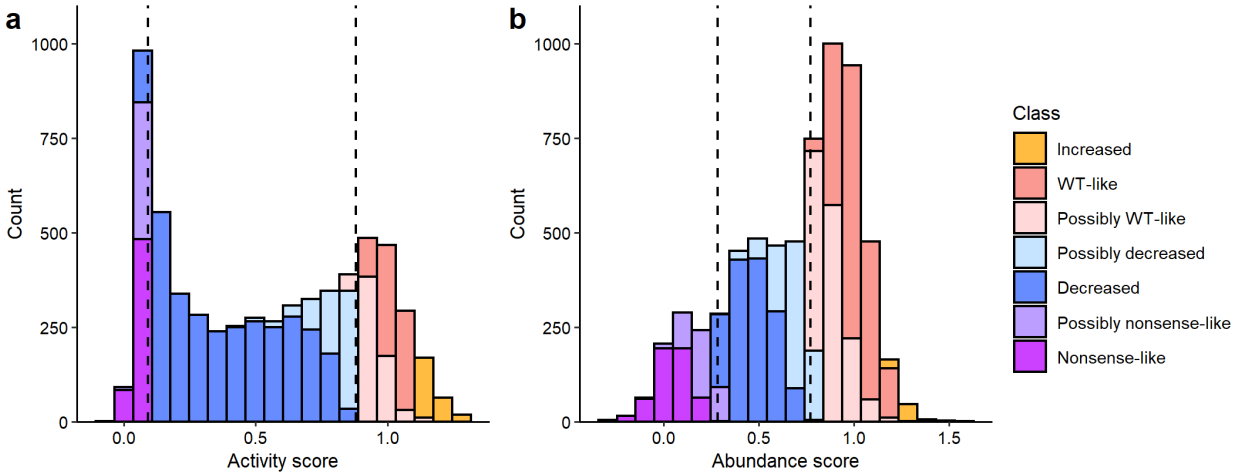


Figure S4. Classification of CYP2C9 scores into classes.
Stacked histograms of CYP2C9 (a) activity and (b) abundance scores categorized into classes. In dotted lines, the 95th percentile of the nonsense distribution (left), and the 5th percentile of the synonymous distribution (right), used for categorization. Variants were categorized by determining whether variant scores and confidence intervals fell within the synonymous and nonsense variant thresholds, as detailed in Material and Methods.

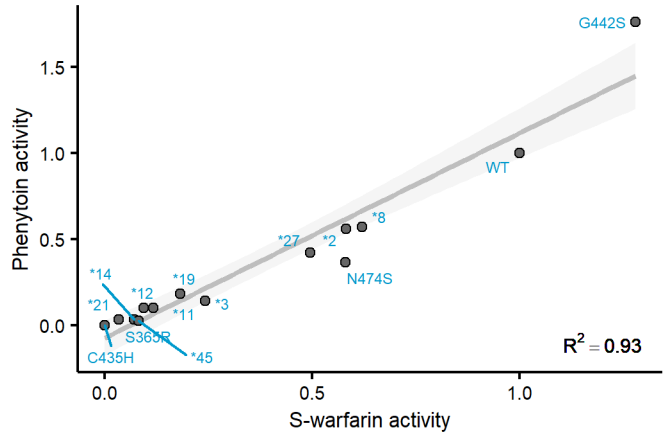


Figure S5. Comparison of gold-standard activity assay with different CYP2C9 substrates.
Individual CYP2C9 alleles were expressed in yeast and microsomes were extracted. The rate of S-warfarin 7-hydroxylation and phenytoin 4-hydroxylation was tested with these microsomes using LC-MS. A scatterplot comparing the CYP2C9 variant activity with these two different substrates is shown. The grey line is the regression line, and shaded area shows the 95% confidence interval. All activities are normalized to wild type rates.

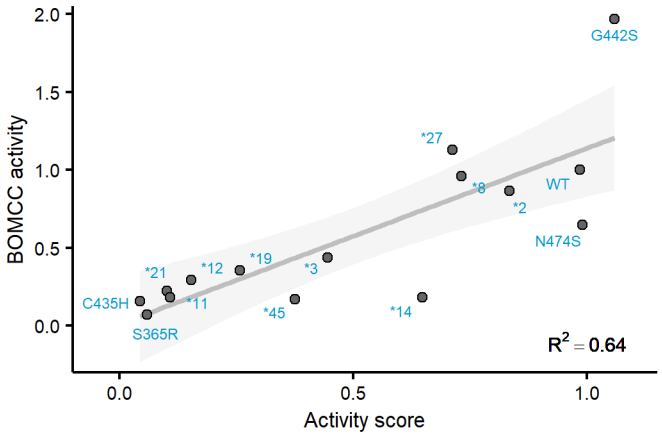


Figure S6. Comparison of CYP2C9 activity scores with fluorogenic activity assay in yeast microsomes.
Scatterplot of CYP2C9 Click-seq activity scores plotted against individually tested CYP2C9 alleles using a fluorogenic substrate. The conversion of BOMCC to CHC (fluorescent) by individual CYP2C9 variants was monitored using a plate reader. The grey line is the regression line, and shaded area shows the 95% confidence interval. All activities are shown normalized to wild type rates.

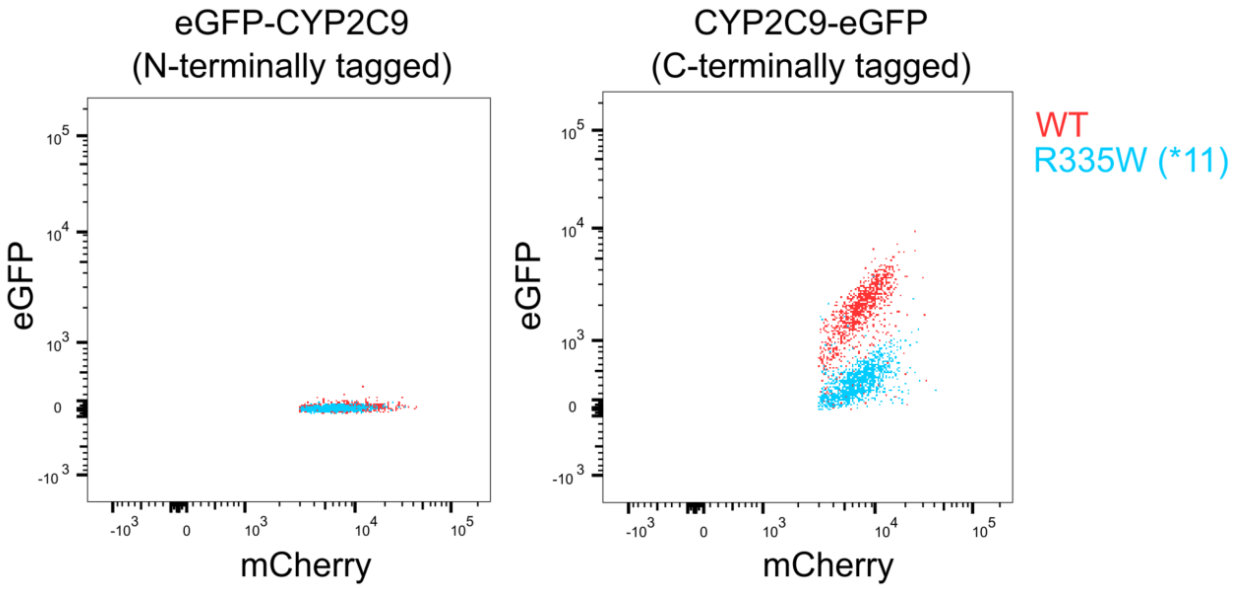


Figure S7. N vs. C-terminal CYP2C9 tagging.
Scatterplots of eGFP vs. mCherry fluorescence for cells expressing either N-terminally eGFP-tagged CYP2C9 (left) or C-terminally eGFP-tagged CYP2C9 (right). WT CYP2C9 shown in red, unstable R335W (*11) variant shown in blue. ~20,000 cells shown each.

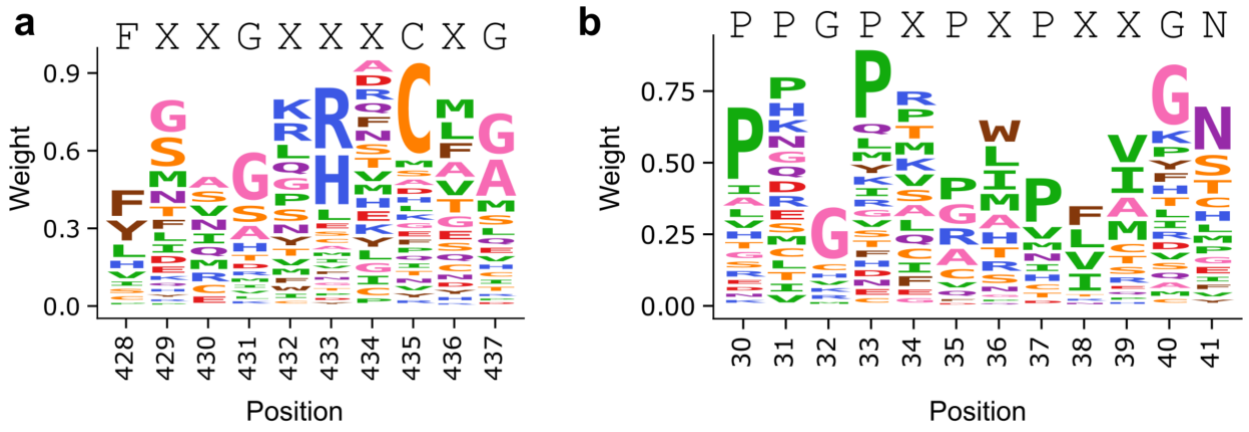


Figure S8. Logo plot of heme binding motif and PPGP motif from DMS data. a) Logo plot of CYP2C9 heme binding motif using Click-seq activity scores, positions 428 to 437. Published heme binding motif[1] shown on top. b) Logo plot of CYP2C9 PPGP linker motif using Click-seq activity scores, positions 30 to 41. Published PPGP motif[2] shown on top. Variant weights calculated by rescaling activity scores from 0 to 1, calculating the frequency of each variant by position, and multiplying frequencies by the fraction of total number of variants present at each position. Variants colored by amino acid type. Figures made using dmslogo (https://github.com/jbloomlab/dmslogo).

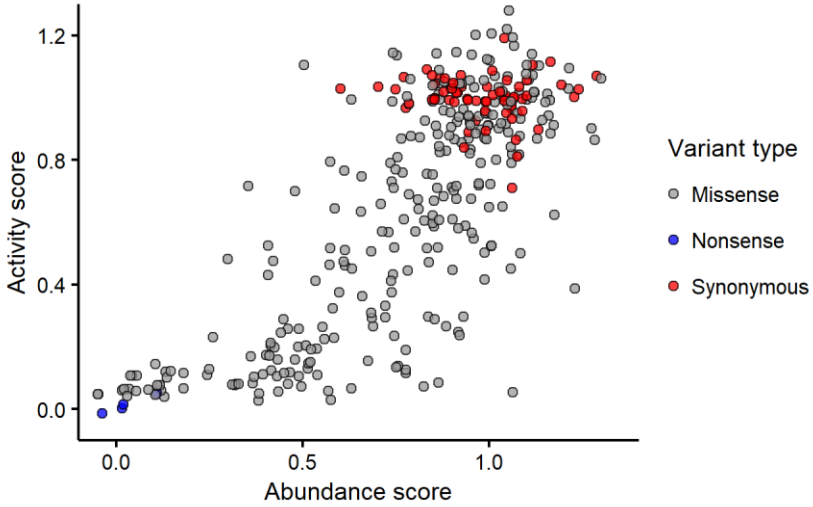


Figure S9. Human CYP2C9 variants with activity and abundance scores. Scatter plot of variant activity vs. abundance score, colored by type of mutation in gnomAD. Variants combined from gnomAD v2 and v3 data, and filtered for missense, stop-gained, and missense variants. A total of 281 gnomAD variants are shown with both activity and abundance scores.

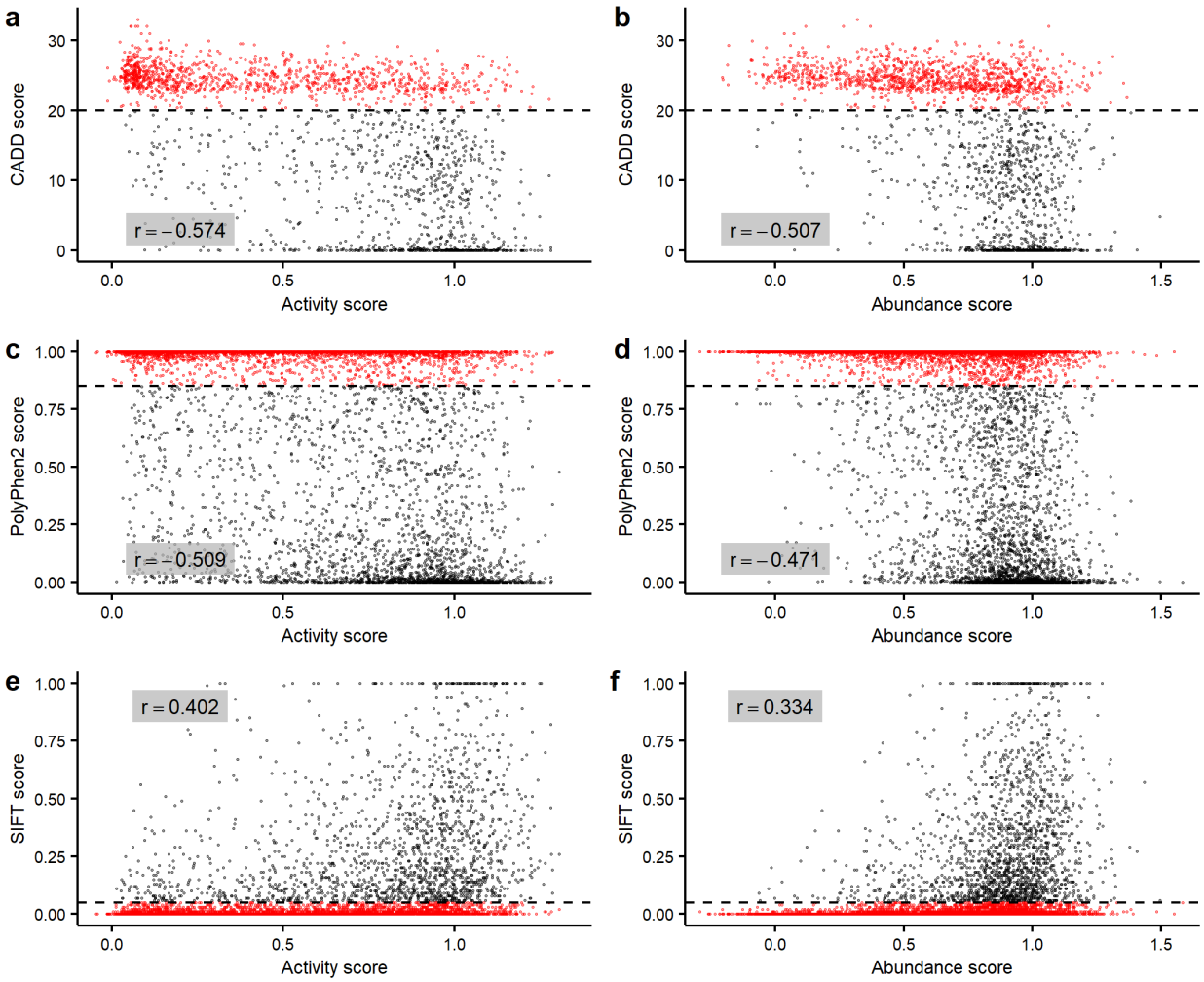


Figure S10. Computational prediction of CYP2C9 missense variant effect. Scatter plots of CYP2C9 missense variant activity score (left) or abundance score (right) vs computational predictions of variant effect. For all plots, correlation (Pearson's r) shown in grey box. In a) and b), CADD score[3] vs activity or abundance score. A CADD score of >20 is considered damaging. Cutoff shown as a dotted line and points shown in red. In c) and d), PolyPhen2 score[4] vs activity or abundance score. A PolyPhen2 score of >0.85 is considered damaging. Cutoff shown as a dotted line and points shown in red. In e) and f), SIFT score[5] vs activity or abundance score. A SIFT score of <0.05 is considered damaging. Cutoff shown as a dotted line and points shown in red.

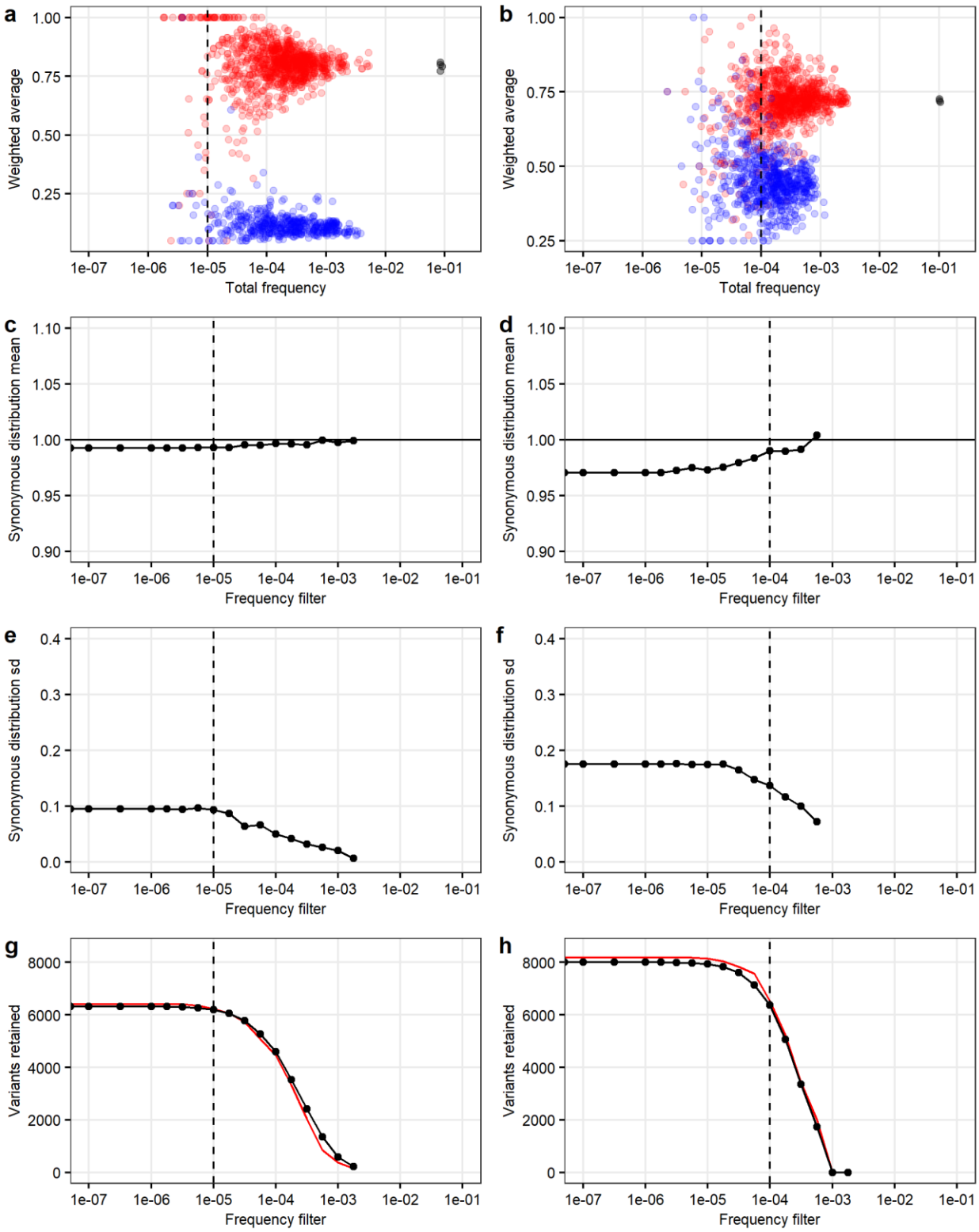


Figure S11. Determining variant frequency filters.
Variant frequency filter used for Click-seq in a,c,e,g) and VAMP-seq in b,d,f,h). In a and b), scatterplots of variant weighted average vs total frequency for wild type (black), synonymous (red), and nonsense (blue) variants, for each of the four or three replicates for the CYP2C9 a) Click-seq and b) VAMP-seq libraries respectively. In c and d), the mean of the synonymous distribution at different frequency filters for the Click-seq and VAMP-seq library, respectively. In e and f), standard deviation 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. For all plots, the frequency filter used for library analysis is shown as a dashed line. For Click-seq this was 10^{-5} (left), and for VAMP-seq this was 10^{-4} (right).

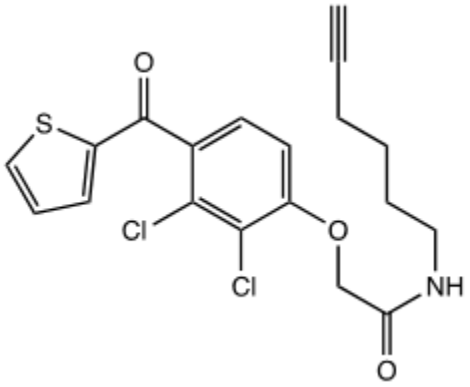
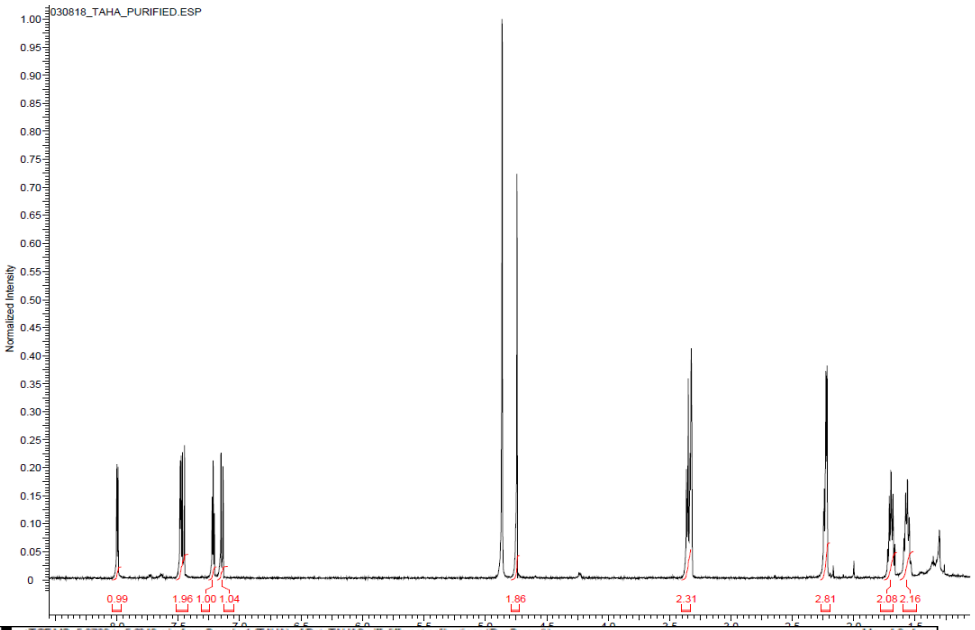
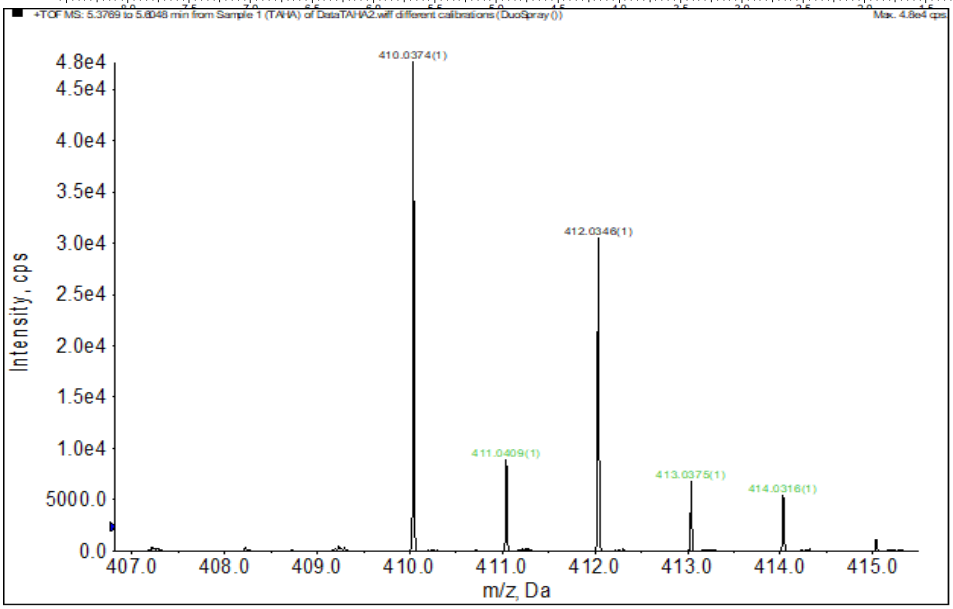


Figure S12. TAHA probe structure.
Chemical structure of tienilic acid hexynyl amide (TAHA).

A


030818_TAHA_PURIFIED.ESP
Normalized Intensity
0.99
1.96
1.00
1.04
1.86
2.31
2.81
2.08
2.16

B


+TOF MS: 5.3769 to 5.6048 min from Sample 1 (TAHA) of DataTAHA2.wiff different calibrations (DuoSpray ())
Max. 4.8e4 cps
410.0374(1)
412.0346(1)
411.0409(1)
413.0375(1)
414.0316(1)
Intensity, cps
m/z, Da

C

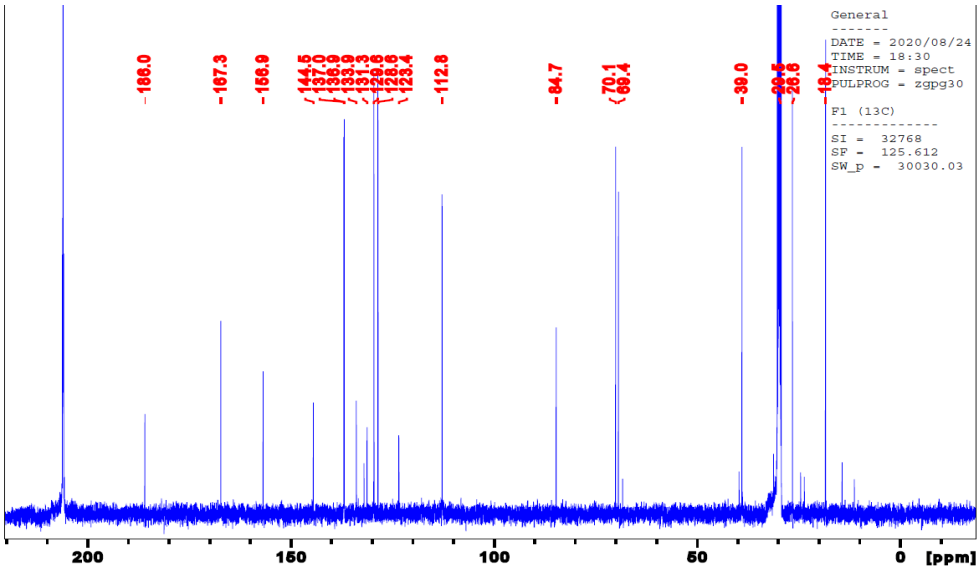

188.0
167.3
166.9
144.5
137.0
136.9
133.9
131.3
129.6
128.6
123.4
112.8
84.7
70.1
68.4
39.0
30.5
26.6
18.4
General
DATE = 2020/08/24
TIME = 18:30
INSTRUM = spect
PULPROG = zgpg30
F1 (13C)
SI = 32768
SF = 125.612
SW_p = 30030.03
200
150
100
50
0
[ppm]

Figure S13. TAHA probe characterization.
(A) ^{1}H spectra of TAHA, (B) ^{13}C NMR spectra of TAHA, and (C) high resolution mass spectra of TAHA depicting isotopic distribution.

SUPPLEMENTAL TABLES

Strain	Genotype
YMD3289	*MATα ura3Δ0 leu2Δ1 his3Δ1 trp1Δ63 HAP1+*
YMD4252	*MATa ura3Δ0 leu2- his3Δ1 trp1Δ63 HAP1+*
YMD4253	*MATα ura3Δ0 leu2Δ1 his3Δ1 trp1Δ63 HAP1+ pep4Δ0 prb1Δ0*
YMD4254	*MATα ura3Δ0 leu2Δ1 his3Δ1 trp1Δ63 HAP1+ pep4Δ0 prb1Δ0 HO:pGAL1-hCPR-FLAG_TRP1*
YMD4255	*MATα ura3Δ0::pGPD-MYC-hb5_URA3 leu2Δ1 his3Δ1 trp1Δ63 HAP1+ pep4Δ0 prb1Δ0 HO::pGAL1-hCPR-FLAG_TRP1*
YMD4256	*MATa ura3Δ0::pGPD-MYC-hb5_URA3 leu2-1 his3Δ1 trp1Δ63 HAP1+ pep4Δ0 prb1Δ0 HO::pGAL1-hCPR-FLAG_TRP1*

Table S1. List of yeast strains generated in this study.
All strains are in a *S. cerevisiae* S288C derivative background.

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 S2. 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 was amplified directly from sorted cells, rather than growing out culture and then amplifying. Asterisk indicates approximate cell numbers.

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 (coverage)	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 majority allele or highest quality read	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 (mean, median barcodes per single amino acid mutation)	38,127 (5.82, 3)	49,015 (5.89, 4)
# single amino acid mutations (percent possible)	**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 S3. Library statistics from barcode-variant mapping.
Barcoded *CYP2C9* libraries sequenced on a Sequel II (Pacific Biosciences). CCS reads (circular consensus reads) generated with ccs2 (Pacific Biosciences).

Additional supplemental files

Table S4. Plasmids and oligos used in this study.

Table S5. CYP2C9 variant activity and abundance scores.

Table S6. CYP2C9 activity and abundance scores by position.

Table S7. CYP2C9 Star Allele CPIC functional annotations.

CYP2C9 star alleles and associate CPIC functional status recommendations. Star allele and associated protein variants taken from pharmvar.org. CPIC functional status and allele evidence level from[6]. CPIC functional status is biochemical functional status

(normal function, decreased function, no function, uncertain function, or unknown function), and evidence level ranges from definitive (strongest), strong, moderate, limited, to inadequate evidence (weakest). Click-seq activity score, activity sd (standard deviation), and activity class (see Methods) shown.

Table S8. CYP2C9 individual variant validation of activity and abundance scores.

Table S9. CYP2C9 optimization data with Click-seq activity-based probes.

SUPPLEMENTAL MATERIAL AND METHODS

Yeast strain engineering

A previously generated S288C derivative strain YMD3289 (*MATα HAP1+ ura3Δ0 leu2Δ1 his3Δ1 trp1Δ63*)[7] was engineered to have improved human P450 activity by increasing protein expression and by expressing human CYP accessory proteins cytochrome P450 reductase (CPR) and cytochrome b5, which are necessary for electron transfer from NADPH to CYP2C9. These accessory proteins are commonly added to yeast P450 expression systems since the homologous CPR and b5 proteins in *S. cerevisiae* do not couple well with mammalian CYP enzymes.[8,9]

First, the vacuolar protease genes *PEP4* and *PRB1* were sequentially knocked out to improve protein expression using the pop-in pop-out method[10] using the vector pRS406[11] with flanking sequences cloned in, resulting in the strain YMD4253. Next, *S. cerevisiae* codon-optimized *POR* sequence (human CPR) (Uniprot: P16435) was synthesized (Integrated DNA Technologies) with a C-terminal FLAG tag (sequence: DYKDDDDK) and cloned into a low-copy p416*GAL1* vector,[12] resulting in the plasmid p416*GAL1-hCPR-FLAG*. The auxotrophic marker *TRP1* was amplified from pRS414[11] and cloned into this vector, and the fragment containing both *GAL1pr*::*hCPR-FLAG* and *TRP1* was amplified, digested with DpnI (NEB R0176), and used to transform the yeast strain YMD4253, resulting in strain YMD4254. Transformants were selected for growth on synthetic media lacking tryptophan (C-trp). Finally, *S. cerevisiae* codon-optimized cytochrome *b5* sequence (Uniprot: P00167) was synthesized (Integrated DNA Technologies) with an N-terminal MYC tag (sequence: EQKLISEEDL) and cloned into a low-copy p416*GPD* vector.[13] A portion of the vector containing both *GPDpr::MYC-hb5* and *URA3* was amplified via PCR, digested with DpnI, and used to transform yeast strain YMD4254, resulting in strain YMD4255. Transformants were selected for growth on synthetic media lacking uracil (C-ura). The fully humanized strain YMD4255 was backcrossed twice with YMD4252, resulting in the strain YMD4256 with genotype *MATa ura3Δ0::GPDpr::MYC-hb5::URA3 leu2Δ1 his3Δ1 trp1Δ63 HAP1+ pep4Δ0 prb1Δ0 ho::GAL1pr::hCPR-FLAG::TRP1* (strain details in Table S1).

The low-copy p41K*GAL1* vector was constructed from the p416*GAL1* vector[12] and the pUG6 vector[14] using Gibson assembly[15] to clone KanMX into p416*GAL1*. *S. cerevisiae* codon-optimized *CYP2C9* sequence (Uniprot: P11712) was synthesized (Integrated DNA Technologies) with a C-terminal HA tag (sequence: YPYDVPDYA) and cloned into p41K*GAL1* using Gibson assembly. Yeast strain YMD4256 was transformed with p41K*GAL1-hCYP2C9-HA* using the standard LiAc protocol referenced above and transformants were selected on YPD media supplemented with 200 µg/mL G418 to maintain the plasmid.

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 spectra was recorded at 25°C in deuterated methanol (CD_3OD) on a 500 MHz Agilent DD2 (Santa Clara, CA) spectrometer, (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). ^{1}H-decoupled ^{13}C NMR ($^{13}C\{^{1}H\}$) spectra was recorded at 25°C in acetone-d_6 (C_3D_6O) on a 500 MHz Bruker Avance DRX-500 (Billerica, MA) spectrometer, equipped with a Bruker triple resonance TXO probehead. Chemical shifts are reported below relative to the solvent peaks in C_3D_6O at 206.7 and 29.9 ppm. $^{13}C\{^{1}H\}$ NMR (125 MHz, C_3D_6O) δ 186.0, 167.3, 156.9, 144.5, 137.0, 136.9, 133.9, 131.3, 129.6, 128.6, 123.4, 112.8, 84.7, 70.1, 69.4, 39.0, 29.5, 26.6, 18.4.

High resolution mass spectrometry (HRMS) was determined via UPLC-MS on a Waters Acquity UPLC (Milford, MA) coupled to an AB Sciex TripleTOF 5600 mass spectrometer (Framingham, MA). Data analysis was performed with AB Sciex Analyst TF 1.7.1. HRMS (ESI+) m/z [M + H] calculated ($C_{19}H_{18}Cl_2NO_3S$) 410.0379, observed 410.0374, δ ppm 1.22. All NMR and mass spectra have been provided in Figure S12 and Figure S13.

Yeast microsomal preparations

A large-scale induction of yeast cells expressing *CYP2C9* wild type 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 18-22 hrs. Cells were pelleted and stored at -80°C until ready for microsome preparations.

Yeast microsomes were prepared as described previously[7,8] 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 M KCl), recovered at 3200 x g, resuspended in 30 mL of TEM buffer (50 mM Tris-HCl, pH 7.4, 1 mM EDTA, 70 mM 2-mercaptoethanol), and incubated at room temperature for 10 minutes. Cells were recovered by centrifugation (3200 x g) 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 (Amsbio) was added. Cells were incubated for ~1 hour at 30°C with agitation until digested. Further steps were performed on ice. Spheroplasts were pelleted at 6732 x g 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 (40-45 W). After 5 minutes on ice, lysed cells were centrifuged for 4 minutes at 1700 x g. The supernatant was then centrifuged at 110,000 x g 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.

BOMCC fluorogenic assay with yeast microsomes

7-Benzyloxymethyloxy-3-cyanocoumarin (BOMCC) (50 μM) was mixed 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: 410 nm, emission: 460 nm, 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 wild type average slope to determine relative BOMCC metabolism.

SUPPLEMENTAL REFERENCES

1. P.B. Danielson, B.S.P. (2002). The Cytochrome P450 Superfamily: Biochemistry, Evolution and Drug Metabolism in Humans. Curr. Drug Metab. *3*, 561–597.
2. Kemper, B. (2004). Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. Toxicol. Appl. Pharmacol. *199*, 305–315.
3. Kircher, M., Witten, D.M., Jain, P., O'roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
4. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
5. Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. *40*, W452–W457.
6. Theken, K.N., Lee, C.R., Gong, L., Caudle, K.E., Formea, C.M., Gaedigk, A., Klein, T.E., Agúndez, J.A.G., and Grosser, T. (2020). Clinical Pharmacogenetics Implementation Consortium Guideline (CPIC) for CYP2C9 and Nonsteroidal Anti-Inflammatory Drugs. Clin. Pharmacol. Ther. *108*, 191–200.
7. 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., et al. (2017). Expression and functional characterization of breast cancer-associated cytochrome P450 4Z1 in saccharomyces cerevisiae. Drug Metab. Dispos. *45*, 1364–1371.
8. Pompon, D., Louerat, B., Bronine, A., and Urban, P. (1996). Yeast expression of animal and plant P450s in optimized redox environments. Methods Enzymol. *272*, 51–64.
9. Hausjell, J., Halbwirth, H., and Spadiut, O. (2018). Recombinant production of eukaryotic cytochrome P450s in microbial cell factories. Biosci. Rep. *38*, 20171290.
10. Dong, J., Wang, G., Zhang, C., Tan, H., Sun, X., Wu, M., and Xiao, D. (2013). A two-step integration method for seamless gene deletion in baker's yeast. Anal. Biochem. *439*, 30–36.
11. Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19–27.
12. Mumberg, D., Muller, R., and Funk, M. (1994). Regulatable promoters of saccharomyces cerevisiae: Comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res. *22*, 5767–5768.
13. Mumberg, D., Müller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene *156*, 119–122.
14. Güldener, U., Heck, S., Fiedler, T., Beinhauer, J., and Hegemann, J.H. (1996). A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res. *24*, 2519–2524.
15. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.