Title

Functional characterization of novel rare CYP2A6 variants and potential implications for clinical outcomes

Journal

Clinical and Translational Science, 15(1)

ISSN

1752-8054

Authors

El-Boraie, Ahmed
Tanner, Julie-Anne
Zhu, Andy ZX
et al.

Publication Date

2022

DOI

10.1111/cts.13135

Peer reviewed

Received: 29 April 2021 | Revised: 23 July 2021 | Accepted: 24 July 2021

DOI: 10.1111/cts.13135

ARTICLE

Functional characterization of novel rare *CYP2A6* variants and potential implications for clinical outcomes

Ahmed El-Boraie[1,2] | Julie-Anne Tanner[3] | Andy Z.X. Zhu[4] | Katrina G. Claw[5] | Bhagwat Prasad[6] | Erin G. Schuetz[7] | Kenneth E. Thummel[8] | Koya Fukunaga[9] | Taisei Mushiroda[9] | Michiaki Kubo[9] | Neal L. Benowitz[10] | Caryn Lerman[11] | Rachel F. Tyndale[1,2,12]

[1]Department of Pharmacology & Toxicology, University of Toronto, Toronto, ON, Canada

[2]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health and Division of Brain and Therapeutics, Toronto, ON, Canada

[3]Myriad Neuroscience, Toronto, ON, Canada

[4]Department of Quantitative Translational Sciences, Takeda Pharmaceuticals, Cambridge, Massachusetts, USA

[5]Division of Biomedical Informatics and Personalized Medicine, University of Colorado, Aurora, Colorado, USA

[6]Department of Pharmaceutical Sciences, Washington State University, Spokane, Washington, USA

[7]Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

[8]Department of Pharmaceutics, University of Washington, Seattle, Washington, USA

[9]Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan

[10]Clinical Pharmacology Research Program, Division of Cardiology, Department of Medicine and Center for Tobacco Control Research and Education, University of California San Francisco, San Francisco, California, USA

[11]Department of Psychiatry, USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California, USA

[12]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

Correspondence
Rachel F. Tyndale, Department of Pharmacology and Toxicology, Room 4326, Medical Sciences Building, 1 King's College Circle, University of Toronto, ON M5S 1A8, Canada.
Email: r.tyndale@utoronto.ca

Funding information
Canada Research Chair in Pharmacogenomics (R.F.T.); Foundation for the National Institutes of Health (NIH) grants PGRN DA020830 (R.F.T. and C.L.), PGRN U01 GM92676 (K.E.T., K.C., and B.P.); National Cancer Institute (NCI) grant R35CA197461 (C.L.); Canadian Institutes of Health Research (CIHR) foundation grant FDN-154294 (R.F.T.) and CIHR project grant PJY-159710 (R.F.T.); the Campbell Family Mental

Abstract

CYP2A6 activity, phenotyped by the nicotine metabolite ratio (NMR), is a predictor of several smoking behaviors, including cessation and smoking-related disease risk. The heritability of the NMR is 60–80%, yet weighted genetic risk scores (wGRSs) based on common variants explain only 30–35%. Rare variants (minor allele frequency <1%) are hypothesized to explain some of this missing heritability. We present two targeted sequencing studies where rare protein-coding variants are functionally characterized in vivo, in silico, and in vitro to examine this hypothesis. In a smoking cessation trial, 1687 individuals were sequenced; characterization measures included the in vivo NMR, in vitro protein expression, and metabolic activity measured from recombinant proteins. In a human liver bank, 312 human liver samples were sequenced; measures included RNA expression, protein expression, and metabolic activity from extracted liver tissue. In total, 38 of 47 rare coding variants identified were novel; characterizations ranged from gain-of-function to

Health Research Institute of the Centre for Addiction and Mental Health (CAMH); the CAMH Foundation; the RIKEN Center for Integrative Medical Science; and the NIH Pharmacogenomics Research Network (PGRN) GM115370.

loss-of-function. On a population level, the portion of NMR variation explained by the rare coding variants was small (~1%). However, upon incorporation, the accuracy of the wGRS was improved for individuals with rare protein-coding variants (i.e., the residuals were reduced), and approximately one-third of these individuals (12/39) were re-assigned from normal to slow metabolizer status. Rare coding variants can alter an individual's CYP2A6 activity; their integration into wGRSs through precise functional characterization is necessary to accurately assess clinical outcomes and achieve precision medicine for all. Investigation into noncoding variants is warranted to further explain the missing heritability in the NMR.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Common *CYP2A6* variants (minor allele frequency >1%) explain 30–35% of the variation in CYP2A6 activity, despite high heritability estimates (60–80%) in the CYP2A6 activity biomarker measure. One hypothesis is that rare coding variants (minor allele frequency <1%) may explain a portion of the missing heritability from pharmacogenes, including *CYP2A6*.

WHAT QUESTION DID THIS STUDY ADDRESS?

What is the relative contribution of rare coding variants in explaining variation in CYP2A6 activity? How necessary is the incorporation of rare coding variants in predicting individual metabolic status, and consequent tailoring of treatment?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

Rare coding variants may explain only a small fraction of the variation on a population level; however, their role may be important on an individual level, altering the predicted metabolic status in a third of the individuals with these rare coding variants.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

Evaluating rare coding variants in pharmacogenes, such as *CYP2A6*, will be valuable in enhancing the investigation of *CYP2A6*'s influence on tobacco addiction and disease pathogenesis, by providing a more accurate reflection of the phenotypic metabolic status through improved genetic assessments.

INTRODUCTION

Nicotine is responsible for tobacco's addictive properties,[1] and the rate of metabolic inactivation is associated with several smoking behaviors and cessation outcomes. Nicotine is predominately metabolized by the cytochrome P450 2A6 (CYP2A6) enzyme to form inactive cotinine (COT). [2] COT is further metabolized to 3-hydroxycotinine (3HC), exclusively by CYP2A6[3]; the 3HC/COT ratio, called the nicotine metabolite ratio (NMR), is an index of CYP2A6 activity and a proxy for nicotine clearance.[3] Variation in the NMR is associated with several smoking phenotypes including acquisition,[4] quantity,[5,6] topography,[7] dependence,[5,6] and smoking-related disease risk,[6,8] and is a biomarker for personalizing smoking cessation treatment.[9-12] For example, in the Pharmacogenetics of Nicotine Addiction and Treatment 2 (PNAT2) smoking cessation trial, normal metabolizers (i.e., higher NMR) quit more on varenicline versus nicotine patch, whereas slow metabolizers (i.e., lower NMR) benefited more from the nicotine patch[10] (equal quit rate, but lower incidence of side effects on the patch).

The NMR requires COT to be at steady-state, and thus it can only be reliably measured in regular smokers. This limits the assessment of the impact of NMR for CYP2A6 substrates[13] as well as for tobacco-related disease consequences (lung cancer, chronic obstructive pulmonary disease [COPD], and diabetes)[6,8] in intermittent-, non-, and former-smokers.

Heritability estimates for the NMR range between 60% and 80%.[14,15] Ninety-six percent of NMR genomewide association studies (GWASs) hits are within or close to *CYP2A6*.[16] Weighted genetic risk scores (wGRSs) based on common variants (minor allele frequency [MAF] >1%) explain 30–35% of the variation in the NMR phenotype.[17,18]

Although substantial, this shortfall is not uncommon; many heritable phenotypes have only a portion of variation accounted for by identified genetic variants (e.g., human height[19]). Several researchers have hypothesized that rare variants (MAF < 1%), which are typically not assessed in GWAS genotyping arrays,[20] contribute to this missing heritability, including for pharmacogenes; *CYP2A6* rare variants could account for up to 40% of the functional variability.[21]

Nevertheless, there are several obstacles in understanding the role of rare variants on CYP function. (1) The high nucleotide sequence similarity/homology within *CYP* subfamily genes creates potential for sequencing errors: sequencing *CYP2A6* is particularly difficult due to its high nucleotide similarity (94%) with the *CYP2A7* pseudogene,[22] making it challenging to verify variants from public sequencing databases. (2) The lack of rare variant functional characterization: often functional assignments are heavily reliant on in silico predictions,[23] which are particularly inaccurate for *CYP* genes.[23] Recently, a combined assortment of in silico prediction tools, weighted specifically for *CYPs*, has been put forth.[23]

In this paper, we present two studies to identify, functionally characterize, and assess the impact of rare *CYP2A6* coding variants. We use two targeted sequencing approaches designed for accurate *CYP2A6* sequencing, in conjunction with variant confirmation through orthogonal genotyping methods. Further, we integrate several in silico, in vivo, and in vitro approaches to characterize variants. Because statistical power is limited when studying rare coding variants individually using solely in vivo measures, we pair our assessments with in vitro functional assays to confirm variant effects on enzyme activity. Our group and others have made use of in vitro cDNA expression systems to assist in interpreting the effect of *CYP* protein-coding variants.[24-26] We assess the extent to which rare coding variants account for heritable variation, and if their incorporation leads to improved prediction of CYP2A6 metabolizer status. Furthermore, we present another set of CYP2A6 measures assessed in a human liver bank,[27] including RNA quantification, protein abundance, and substrate metabolic activity assessed using human liver microsomes (HLMs).

METHODS

CYP2A6 read alignment simulation

A simulation was performed, aligning simulated reference *CYP2A6* sequencing reads of varying lengths and insert sizes to assess public whole-genome sequencing accuracy, where similar analyses have been described for the *CYP2D6* gene.[28] The read alignment simulation suggests there are several exons in which *CYP2A6* reads could misalign, specifically exons 5 and 9 (Figure S1). Further details are described in the Supplementary Material.

The following studies were approved by institutional review boards at all participating sites. Flowcharts of the different analyses and cohorts are summarized in (Figure S2).

Smoking cessation trial

Study population

There were 1684 treatment-seeking smokers, including 541 of self-reported African ancestry (AFR) and 1026 of self-reported European ancestry (EUR), were screened at baseline as part of the PNAT2 (NCT01314001) smoking cessation trial.[10] The remaining 117 subjects consisted of small numbers of other racial/ethnic populations; thus we focused on individuals of AFR and EUR ancestry.

Sequencing

Targeted deep exon sequencing was performed through the Illumina MiSeq sequencing system on the *CYP2A6* gene.[18] All exons were amplified through multiplex polymerase chain reaction (PCR) using gene-specific primers, where amplicon sizes ranged from 260 to 490 bp; where the difficult-to-sequence exon 9 was captured through a nested PCR approach. Paired-end sequencing reads were mapped to the Hg19 reference genome using BWA-mem (version 0.7.15).[29] Variants were called using GATK's haplotype caller pipeline (version 3.3).[30] Validation of the sequencing protocol was assessed in a sample of 120 polymerase chain reaction Japanese individuals (not part of the trial) to ensure targeted sequencing accuracy of *CYP2A6* compared to Sanger sequencing, details of this validation can be found in Supplementary Table S1. All coding variants not yet described as a *CYP2A6* * allele (pharmvar.org) were considered novel (i.e., whereas variants may be catalogued in other public sequencing databases), they have not previously been functionally characterized. Furthermore, all novel variants identified were confirmed through Sanger sequencing, the gold-standard for validating next-generation sequencing variant calls.[31]

In vivo measure quantification

Free (unconjugated) concentrations of COT and 3HC, used to calculate the NMR, were measured from whole blood, where samples were collected at intake while individuals were smoking ad libitum.[32]

In silico prediction

To designate in silico assignments, an optimized pharmacogenomic/absorption, distribution, metabolism, and excretion (ADME) prediction framework, published elsewhere,[23] was utilized. Details are described in the Supplementary Material.

In vitro functional characterization

Details on the construction, expression, measure of protein quantity, and enzyme activity for the *CYP2A6* variant constructs has been described elsewhere[24,25] and is described in full in the Supplementary Material. In vitro metabolism of nicotine to cotinine was used to determine CYP2A6 activity for each construct, as described elsewhere[24,25] and in the Supplementary Material.

Variant functional assignments

Assignments of loss-of-function, decrease-of-function, neutral, and gain-of-function were determined relative to the reference constructs *CYP2A6* wildtype (WT) and *CYP2A6*17* (known decrease-of-function variant). An aggregate functional assignment based on the three measures was made, and the level of evidence was assigned based on agreement between measures (3-most confident, 1-least confident). In cases where the in vitro and in vivo measures were in disagreement, the construct would be designated "inconsistent evidence" (IE) and assigned a level of evidence of 1. Together, two versions of aggregate functional assignments were created, a four-level assignment distinguishing between loss and decrease-of-function, and a simplified three-level assignment, which merged these two groups into a single decrease-of-function group. In vitro assignments for known rare coding variants were based on previous publications.[24,26,33]

Human liver bank

Study population

There were 312 human liver tissue samples, including 298 EUR individuals from two liver banks (smoking status unknown), (1) the St. Jude Liver Resource at the St. Jude Children's Research Hospital (Memphis, TN) and (2) the University of Washington Human Liver Bank (Seattle, WA).[27] The remaining 14 subjects consisted of small numbers of other racial/ethnic populations.

Sequencing and in vitro measures

CYP2A6 DNA sequencing, mRNA and protein quantification, and enzyme activity assays have been described previously,[27,34] and is briefly redescribed in the Supplementary Materials. All novel variants identified were confirmed through RNA sequencing using a different set of primers and library construction than those in the DNA sequencing.

Statistical analyses

Rare variant analyses

Rare variant association testing using the SKAT method was performed in BioBin version 2.3.0,[35] where variants were binned based on an MAF cutoff of less than 1% and restricted to comparing *CYP2A6* coding variants to the reference groups.

Reference groups

Reference groups were defined without a *CYP2A6* rare variant and without one of the common *CYP2A6* * alleles that are associated with the NMR (and included in their respective wGRSs). For AFR individuals, excluded common alleles were *CYP2A6 *1X2*, *4, *9, *12, *17, *20, *25, *26, *27, and *35*.[18] For EUR individuals, excluded common alleles were *CYP2A6 *2, *4, *9, and *12*.[17]

Updated wGRSs

CYP2A6 activity wGRSs have been described for both AFR[18] and EUR[17] populations, based on 11 and seven common variants, respectively. For the wGRS analyses, ancestry was determined using principal components analysis, restricting to 504 AFR and 933 EUR individuals from the PNAT2 trial. To incorporate rare coding variants, effect sizes for the rare coding variants group were calculated through SNPTEST (version 2.5.2), as described previously[17]; remaining statistical analyses were performed in R (version 3.6.0) or RStudio (version 1.1.463). Effect sizes were computed either from the novel and previously known rare coding variants grouped together, or the novel rare coding variants alone. The variant weight was estimated by multiplying the standard deviation of NMR (0.181 and 0.205 in the AFR[18] and EUR[17] groups, respectively) by the effect size (i.e., beta) of the grouped rare variant alleles. The NMR was not normally distributed (by the Shapiro-Wilk test) and was therefore log-transformed, which best represents the nicotine clearance rate.[17] Linear regression assessed log-transformed NMR

(log-NMR) variation accounted for by the wGRS models, and residuals were calculated for individual participants from the wGRS line of best fit (i.e., distance between the wGRS model and outlier data points). A paired *t*-test was used to assess the improvement in residual distribution after incorporating rare coding variants into the wGRS.

RESULTS

Smoking cessation trial

Identification of rare protein-coding variants in CYP2A6 among 541 of African ancestry and 1026 individuals of European in the PNAT2 trial

There were 37 rare coding variants identified across 70 individuals (Table 1) through targeted *CYP2A6* exon sequencing. Of the 37, eight were known, whereas 29 were novel and have not been previously identified as part of a *CYP2A6* * allele or functionally characterized (pharmvar.org). All novel variants would be defined as rare (MAF <1%) based on the sequenced sample and based on public sequencing databases (e.g., gnomAD). Some of the novel variants were found concurrently with other novel variants or other common variants (Table S2). There were 15 rare coding variants identified in 31 of 541 AFR individuals, and 23 rare coding variants in 39 of 1026 EUR individuals (one of these variants overlapped between the AFR and EUR groups). There was a higher frequency of rare coding variants overall in AFRs (31/541; 5.7%) versus EURs (39/1026; 3.8%).

In vivo associations

When grouped, rare coding variants were associated with decreased NMR relative to the reference group (those without rare coding variants and without common functionally relevant *CYP2A6* * alleles; Figure 1). However, relative to the reference group, some variants appeared to have neutral or gain-of-function relationships with the NMR; these were more common in the EUR smoker group than in the AFR smoker group.

In silico predictions

The in silico-based predictions using the optimized ADME-prediction framework to assess variant outcomes were not able to distinguish the unique in vivo NMR associations between variants (Figure S3), which prompted further testing using in vitro approaches.

In vitro functional characterization

All 29 novel rare coding variants were introduced into respective *CYP2A6-POR* bicistronic constructs and proteins were expressed. A WT, *CYP2A6*17* construct (a common decrease-of-function *CYP2A6* variant), and a construct with *CYP2A6* excised (CYP-DEL) were expressed concurrently with the novel variants and used as reference points.[24-26] For immunoblotting, quality control checks were performed to confirm antibody specificity and protein identification in the WT and CYP-DEL constructs, and a commercially expressed CYP2A6 protein source against pooled HLM where CYP2A6 and POR is expressed; further details are explained in Supplementary Figure S4. CYP2A6 protein levels from the constructs were determined using a standard curve of commercially expressed CYP2A6 protein (250–1000 fmol; Figure S5).

The ratio of CYP2A6 protein levels to POR protein levels was determined by loading approximately equivalent POR (internal plasmid expression control; Figure 2a). The sample dilution curves (Figure S5) were used to determine an equivalent amount of POR. The ratio of CYP2A6 to POR was used to evaluate changes in CYP2A6 expression for each variant relative to the WT construct (Figure 2b). The in vitro (i.e., 2A6/POR protein expression and V_{max}/K_m catalytic efficiency corrected for CYP2A6 expression level) and in vivo (i.e., NMR) measures are summarized in (Tables 2 and 3).

Variant integration into CYP2A6 wGRSs

Most individuals with a functionally important rare variant (variants with a level of evidence of 2 or 3; Table 3) were outliers from the lines of best fit in the original wGRS to log-NMR correlations, as assessed by their residual value (Figure 3a,b; i.e., distance between the wGRS line of best fit and data points, to measure of how well the line fits for individuals sequenced with rare coding variants). In the AFR population, the average residual value of individuals with a functionally relevant rare variant was 0.238, compared to the general AFR population with an average residual of 0.188. In other words, the wGRS (based on common variants) is a poorer fit for those with rare coding variants than for everyone else. Likewise, in the EUR population, these values were 0.213 and 0.140, respectively. This suggests that the wGRS score assigned for those with rare coding variants was less reflective of their log-NMR (i.e., weaker metabolizer status prediction).

The rare coding variants were then weighted as a collective decrease-of-function group (based on the 3-level functional assignment) and incorporated into the previously reported CYP2A6 wGRSs. Individuals without rare

TABLE 1 Rare *CYP2A6* variants identified in study one

Amino acid change	rsID	hg19 Position	hg19 REF	hg19 ALT	Transcript change	Frequency count (ancestry)	Novel or known	In silico prediction
I61F	rs200554095	41355885	T	A	c.181A>T	1 (AFR)	Novel	0 = Neutral
R128L	rs4986891	41354629	C	A	c.383G>T	4 (AFR)	*CYP2A6*26*	1 = Deleterious
S131A	rs59552350	41354621	A	C	c.391T>G	4 (AFR)	*CYP2A6*26*	1 = Deleterious
R190L	rs571335587	41354209	C	A	c.569G>T	1 (AFR)	Novel	0.8 = Deleterious
Q239K	rs138978736	41352896	G	T	c.715C>A	4 (AFR)	Novel	0 = Neutral
P264T	rs761666827	41352821	G	T	c.790C>A	1 (AFR)	Novel	0.8 = Deleterious
R265Q	rs140471703	41352817	C	T	c.794G>A	3 (AFR)	*CYP2A6*41*	1 = Deleterious
M275I	rs111869995	41352786	C	A	c.825G>T	1 (AFR)	Novel	0.6 = Deleterious
E330D	rs137904044	41351370	C	A	c.990G>T	1 (AFR)	Novel	0.8 = Deleterious
R333X	rs61605570	41351363	T	A	c.997A>T	3 (AFR)	Novel	1 = Deleterious
M368I	rs772964366	41351256	C	G	c.1104G>C	1(AFR)	Novel	0.2 = Neutral
T378I	rs114558780	41351227	G	A	c.1133C>T	3 (AFR)	Novel	0.4 = Neutral
E419K	rs768416963	41350584	C	T	c.1255G>A	1 (AFR)	Novel	0 = Neutral
I434V	rs1302192284	41350539	T	C	c.1300A>G	2 (AFR)	Novel	0 = Neutral
R64C	rs374515279	41355875	G	A	c.191G>A	2 (1 AFR, 1 EUR)	Novel	0 = Neutral
P35L	rs377713545	41356228	G	A	c.104C>T	1 (EUR)	Novel	1 = Deleterious
R64H	rs374515279	41355875	C	T	c.191G>A	1 (EUR)	Novel	0.4 = Neutral
E97K	rs145308399	41355777	C	T	c.289G>A	8 (EUR)	Novel	0.6 = Deleterious
G115D	rs758479488	41354668	C	T	c.344G>A	1 (EUR)	Novel	1 = Deleterious
V140G	rs777098658	41354593	A	C	c.419T>G	1 (EUR)	Novel	0.6 = Deleterious
V140A	rs777098658	41354593	A	G	c.419T>C	1 (EUR)	Novel	0.6 = Deleterious
D158E	rs60605885	41354538	G	C	c.474C>G	3 (EUR)	*CYP2A6*22*	0 = Neutral
L160I	rs60563539	41354534	G	T	c.478C>A	3 (EUR)	*CYP2A6*22*	0 = Neutral
K194E	rs199916117	41354198	T	C	c.580A>G	1 (EUR)	*CYP2A6*15*	0 = Neutral
R257C	rs145157460	41352842	G	C	c.769C>T	1 (EUR)	Novel	0.4 = Neutral

(Continues)

TABLE 1 (Continued)

Amino acid change	rsID	hg19 Position	hg19 REF	hg19 ALT	Transcript change	Frequency count (ancestry)	Novel or known	In silico prediction
R257G	rs145157460	41352842	G	A	c.769C>G	2 (EUR)	Novel	0 = Neutral
E279Q	rs58261757	41351999	C	G	c.835G>C	1 (EUR)	Novel	0.2 = Neutral
I300T	rs148693084	41351935	A	G	c.899T>C	2 (EUR)	Novel	0 = Neutral
R311C	rs58571639	41351903	G	A	c.931C>T	1 (EUR+AFR)[a]	Novel	0.8 = Deleterious
YGFL312-315L	Unavailable	41351891	AAGAAGCCAT	A	c.944del	1 (EUR)	Novel	1 = Deleterious
V323M	rs1303839356	41351867	C	T	c.967G>A	1 (EUR)	Novel	0.4 = Neutral
R339Q	rs150247689	41351344	C	T	c.1016G>A	1 (EUR)	Novel	0.8 = Deleterious
M352T	rs143841823	41351305	A	G	c.1055T>C	3 (EUR)	Novel	0.2 = Neutral
F362S	rs778019189	41351275	A	G	c.1085T>C	1 (EUR)	Novel	1 = Deleterious
S465T	rs746173331	41349793	A	T	c.1393T>A	1 (EUR)	Novel	0 = Neutral
I471T	rs5031016	41349774	A	G	c.1412T>C	1 (EUR)	*CYP2A6*7/*19*	0.6 = Deleterious
G479V	rs5031017	41349750	C	A	c.1436G>T	2 (EUR)	*CYP2A6*5*	1 = Deleterious

Note:: In silico prediction based off framework using several prediction tools that are optimized for ADME or pharmacogenes.[23] Novel is defined as not previously functionally characterized or part of a known *CYP2A6** allele. X: Premature stop codon.

Abbreviations: ADME, absorption, distribution, metabolism, and excretion; AFR, African ancestry; ALT, alternate allele on the positive strand of the hg19 reference genome build; EUR, European ancestry; REF, reference allele.

[a]Participant self-reported race is multi-racial.

coding variants, or with rare coding variants deemed to have inconsistent evidence (Table 3), retained their original wGRS values. The overall beta evaluated for the decrease of function rare variant group in the AFR population was −0.695 (the consequent variant weight was −0.126), and this beta was consistent when incorporating the list of known rare coding variants. Likewise, in the EUR population, the overall beta was −0.819 (variant weight −0.168).

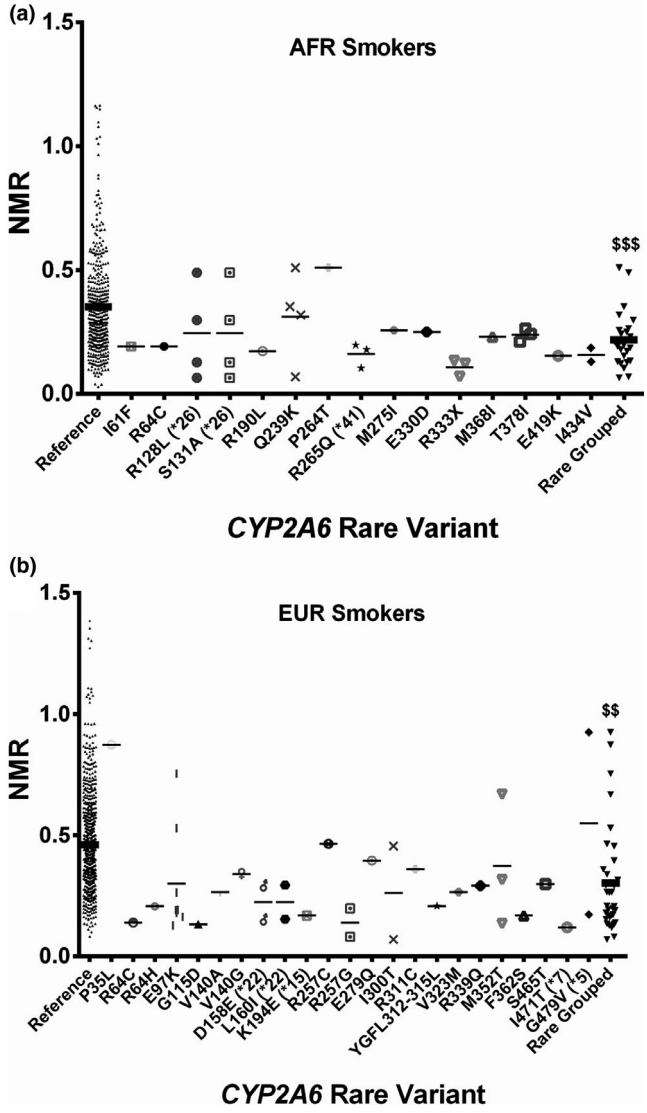


FIGURE 1 Beeswarm plots of the rare coding variants plotted individually, and grouped, with their NMR values in the (a) AFR ($N = 541$) and (b) EUR ($N = 1026$) populations. Reference: Individuals without a *CYP2A6* rare variant and without any common functional *CYP2A6* * alleles. X: Premature stop Codon. Statistical analyses were based off the SKAT test statistic for continuous traits comparing the Rare Grouped and Reference groups. $$: $p < 0.001$ $$$: $p < 0.0001$. AFR, African-ancestry; EUR, European-ancestry; NMR, nicotine metabolite ratio

There was only a minor improvement in the overall variance captured, as expected, given the small number of people with functionally important rare coding variants. In AFR smokers, the variance of NMR captured increased from $R^2 = 30.7\%$ (Figure 3a) to 31.9% (Figure 3c), whereas in EUR smokers, the variance increased from $R^2 = 33.8\%$ (Figure 3b) to 34.8% (Figure 3d). Similar findings were observed when incorporating known demographic covariates of the NMR (sex, age, and body mass index [BMI]) in both AFR smokers ($R^2 = 33.5$–34.6%) and EUR smokers ($R^2 = 37.6$–38.6%). However, when examining the specific individuals with functionally important rare coding variants in the AFR population ($N = 18$), the average residual value for individuals with a rare variant decreased from 0.238 to 0.158. Likewise, in the EUR population ($N = 21$), the average residual value for individuals with a rare variant decreased from 0.213 to 0.179. Together, there was a significant improvement in the residuals for those with rare coding variants ($p < 0.001$ based on a paired *t*-test).

In the AFR group, eight individuals were reclassified from normal to slow metabolizer status after incorporating the rare coding variants (Figure 3c). In the EUR group, four individuals were reclassified from normal to slow metabolizer status after incorporating the rare coding variants (Figure 3d). For these 12 individuals, this reclassification based on their improved wGRS predicted metabolizer status (based on a 2.089 AFR and 2.14 EUR cutoff point)[17,18] lead to improved concordance with their NMR-based metabolizer status (based on a 0.31 cutoff point).[10]

Human liver bank

Identification and characterization of rare protein-coding variants in CYP2A6 among 312 individuals in a human liver bank

Ten rare coding variants (one known and nine novel) were identified among 17 individuals of the 312 total individuals sequenced (Table S3). When grouped, the rare coding variants were not associated with overall changes in CYP2A6 mRNA or protein (Figure 4a,b). However, they were overall associated with a decrease in enzyme activity (Figure 4c), suggesting some variants may alter transcription or translation/stability and resulting activity, whereas most coding variants directly altered intrinsic metabolic activity. Like in the Smoking Cessation Trial, the variants demonstrated a spectrum of functional assignments. Two of the variants identified in the Smoking Cessation Trial were also identified in the Human Liver Bank (E97K and R311C); the effects on CYP2A6 activity were consistent

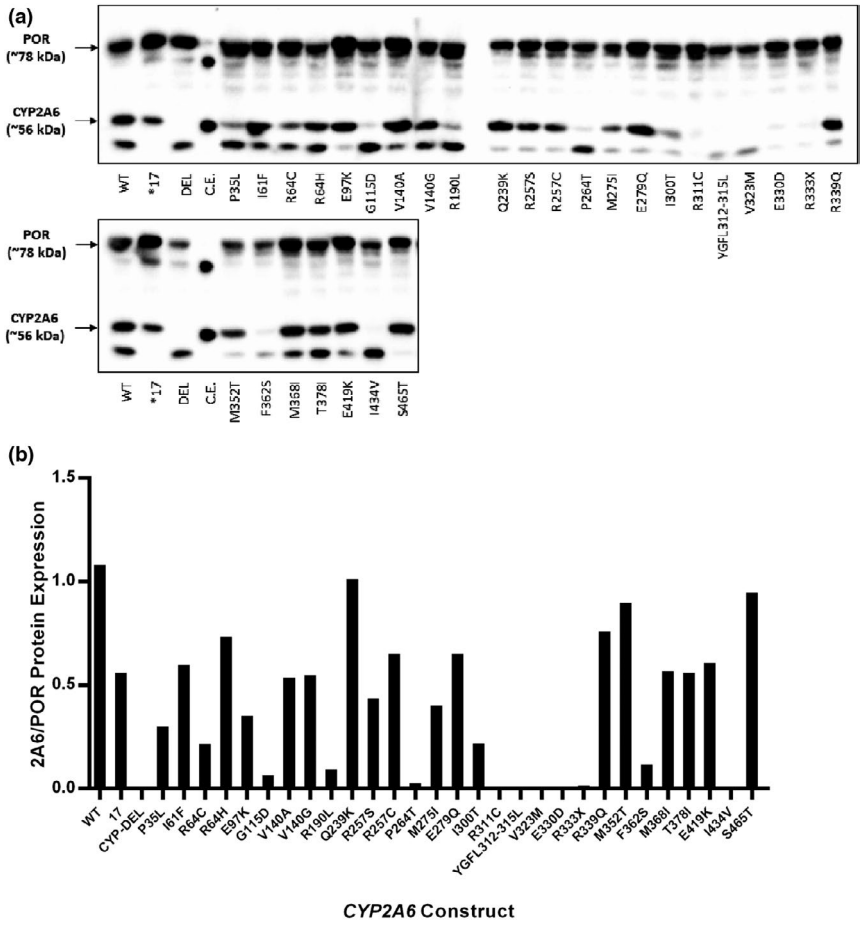


FIGURE 2 (a) Western blot loading approximately equal POR amounts for each novel CYP2A6 rare variant identified in Smoking Cessation Trial. Sample loading was based on sample dilution curves (Figure S4), adjusting to achieve similar POR amounts between constructs (i.e., more sample was loaded for variant constructs with low expression), and within the respective POR and 2A6 linear ranges. (b) Bar graph displaying the CYP2A6/POR protein expression ratios for each variant construct. kDa, kilodalton; WT, wildtype expressed in *E. coli*; DEL, CYP2A6 excised from bicistronic construct with POR intact; C.E., Commercially Expressed CYP2A6 (CAT #456254; Corning); X, Premature stop Codon

for E97K between the two studies. The 311C variant was associated with higher mRNA and protein expression in the human liver bank, neutral in vivo activity (NMR) in the smoking cessation trial and decreased intrinsic in vitro activity (based on the constructs); overall, this suggests that R311C is associated with higher expression, but lower intrinsic enzyme activity, leading to neutral in vivo activity. A summary of the functional characterizations of the rare coding variants identified in the Human Liver Bank are in (Figure 4d).

DISCUSSION

There were 47 rare coding variants identified among 87 individuals of 1996 sequenced in the two studies, 38 of which were novel and have not been previously functionally characterized. These rare coding variants were collectively associated with decreased enzyme activity, as observed in vivo by the NMR (Figure 1) and in vitro through metabolic assays using cDNA expressed enzymes (Tables 2 and 3) and HLM (Figure 4c). Most of these coding rare variants impose effects on metabolic activity (Tables 2 and 3, and Figure 4); some mediate these effects through changes in mRNA/protein expression (Figure 4), but most altered intrinsic enzyme properties, perhaps through alterations to the enzyme's access channel or affecting substrate binding. Most rare coding variants reported in *CYP* genes, including previously for *CYP2A6*, are associated with decreased CYP enzyme activity[24,36]; exploration of noncoding rare variants may yield more neutral or gain of function effects. The region containing the stretch of amino acid residues 300–330 may play an important function in the protein folding and stability of CYP2A6, as indicated by a consistent reduction of CYP2A6 protein from variant constructs in this region (Figure 2), and the high conservation of this region according to in silico predictions (Table 1). At a population level, the impact of rare coding variants appears to be small, represented by a minor increase in the variance (R^2) captured in the NMR (Figure 3c,d; i.e., ~ 1% of variance was explained by the rare coding variants in the AFR and EUR samples). Even when controlling for nongenetic factors (sex, age, and BMI), the variance captured by the rare variants was still 1%, suggesting that this fraction of variation is directly explained by the variants and is not skewed by nongenetic factors.

TABLE 2 Detailed V_{max}, K_m, V_{max}/K_m (Turnover), and percentage of WT (Turnover Rates) for all novel rare variants identified in the Smoking Cessation Trial

Variant construct	V_{max} (pmol COT/min/pmol 2A6)	K_m (μM)	V_{max}/K_m (nl/min/pmol 2A6)	% WT (turnover rate)
WT	6.0	48.0	125	100
*V365M (CYP2A6*17)*	3.3	42.2	79	64
P35L	6.0	134.0	45	36
I61F	8.7	63.5	137	110
R64H	5.6	60.6	93	74
R64C	3.7	39.8	93	75
E97K	7.8	82.9	95	76
Q239K	12.7	67.2	189	152
V140A	5.7	35.7	160	128
V140G	3.5	34.9	100	80
R190L	2.9	153.4	19	15
R257C	5.0	124.3	40	32
R257G	5.4	152.7	35	28
P264T	3.1	87.4	36	29
M275I	7.3	68.6	106	85
E279Q	4.6	62.2	74	60
I300T	7.2	185.3	39	31
R311C	2.8	107.1	26	21
YGFL312-315L	1.0	112.0	9	7
V323M	1.1	130.2	8	7
E330D	8.9	97.1	91	73
R333X	0.5	375.9	1	1
R339Q	8.2	41.0	201	161
M352T	11.2	55.5	201	161
F362S	3.4	108.5	31	25
M368I	10.3	74.1	140	112
T378I	4.9	111.3	44	35
E419K	3.9	75.5	52	42
I434V	3.5	83.0	42	34

Abbreviations: COT, cotinine; K_m, kinetic metabolite; WT, wildtype; V_{max}, maximum value; X, premature stop Codon.

However, on an individual level, the inclusion of rare coding variants led to a shift for these individuals from predicted normal to predicted slow metabolizer status (Figure 3c,d), which was more in line with their NMR metabolizer status (i.e., 12 of 39 individuals sequenced with rare coding variants were re-assigned to slow metabolizer status when incorporating rare coding variants into their wGRS). In the context of smoking cessation treatment, based on the findings from the PNAT2 clinical trial,[10] to derive the greatest quitting rates with the fewest incidences of side effects, treatment recommendations for 12 of 39 individuals would shift from varenicline to the nicotine patch. The importance of integrating rare coding variants for these individuals was also observed by the significant reduction in the residual values from the wGRS line of best fit (i.e., the wGRSs now worked more accurately for those with rare coding variants).

The 1% of variation captured falls short of the 40% estimate previously suggested to explain functional variation in CYP2A6.[21] Three potential elements may have contributed to the high estimate. The first is the use of public sequencing databases. As demonstrated by the read alignment simulation, there is a potential for sequencing errors if the sequencing protocol does not meet rigorous criteria in terms of read length or selection of probes that prevent off-target sequencing (Figure S1); caution is necessary when assessing whole-genome sequencing databases. The second element is the dependence on in

TABLE 3 Summary of in vitro assessments, in vivo associations, and variant construct functional assignments in AFR ($N = 541$) and EUR ($N = 1026$) populations

	Variant construct	CYP2A6/POR protein		V_{max}/K_m (nl/min/pmol 2A6)		Average NMR (Individual NMR)[a]		Aggregate functional assignment (4-level)	Aggregate functional assignment (3-level)	Level of evidence
AFR										
Reference	WT	1.08	N	125	N	0.351	N	N	N	—
	V365M (*CYP2A6*17*)	0.56	D	79	D	0.235	D	D	D	—
Novel rare	R64C	0.22	L	93	D	0.192	L	L	D	3
	R190L	0.09	L	19	L	0.173	L	L	D	3
	R333X	0.01	L	1	L	0.108 (0.069, 0.123, 0.133)	L	L	D	3
	E419K	0.61	D	52	L	0.154	L	L	D	3
	I434V	0.09	L	42	L	0.158 (0.13, 0.186)	L	L	D	3
	M275I	0.40	D	106	D	0.257	D	D	D	3
	E330D	0.01	L	92	D	0.250	D	D	D	3
	T378I	0.56	D	44	L	0.239 (0.211, 0.242, 0.264)	D	D	D	3
	I61F	0.60	D	137	N	0.192	L	D	D	2
	M368I	0.57	D	140	N	0.231	D	D	D	2
	Q239K	1.01	N	189	G	0.394 (0.319, 0.353, 0.509)	N	N	N	3
	P264T	1.01	N	36	L	0.509	G	N	N	2
Known rare	F118L, R128L, S131A (*CYP2A6*26*)[a]	0[c]	L	0%[d]	L	0.245 (0.065, 0.128, 0.298, 0.489)	D	L	D	3
	R265Q (*CYP2A6*41*)	0[e]	L	5%[d]	L	0.16 (0.104, 0.179, 0.197)	L	L	D	3
EUR										
Reference	WT	1.08	N	125	N	0.458	N	N	N	—

(Continues)

TABLE 3 (Continued)

	Variant construct	CYP2A6/POR protein		V_{max}/K_m (nl/min/pmol 2A6)		Average NMR (Individual NMR)[a]		Aggregate functional assignment (4-level)	Aggregate functional assignment (3-level)	Level of evidence
Novel rare	R64C	0.22	L	93	D	0.139	L	L	D	3
	G115D	0.06	L	9	L	0.132	L	L	D	3
	R257S	0.43	D	35	L	0.139 (0.08, 0.197)	L	L	D	3
	I300T	0.22	L	39	L	0.262 (0.069, 0.455)	D	L	D	3
	YGFL312—315L	0.00	L	9	L	0.207	D	L	D	3
	V323M	0.00	L	8	L	0.265	D	L	D	3
	F362S	0.12	L	32	L	0.168	L	L	D	3
	R64H	0.73	D	93	D	0.206	D	D	D	3
	E97K	0.35	L	95	D	0.29975 (0.127, 0.162, 0.18, 0.191, 0.192, 0.263, 0.529, 0.754)	D	D	D	3
	V140G	0.55	D	100	D	0.339	D	D	D	3
	V140A	0.54	D	160		0.265	D	D	D	2
	M352T	0.90	N	201	G	0.372 (0.134, 0.315, 0.668)	N	N	N	3
	R339Q	0.76	D	201	G	0.291	D	N	N	2
	P35L	0.30	L	45	L	0.873	G	I.E.	I.E.	1
	R257C	0.65	D	40	L	0.464	N	I.E.	I.E.	1
	E279Q	0.65	D	74	D	0.395	N	I.E.	I.E.	1
	R311C	0.00	L	26	L	0.359	N	I.E.	I.E.	1
	S465T	0.95	N	189	G	0.298	D	I.E.	I.E.	1
Known rare	L160I, D158E (*CYP2A6*22*)[a]	Unknown		66%[d]	D	0.223 (0.153, 0.293)	D	D	D	2
	K194E (*CYP2A6*15*)	Unknown		219%[d]	G	0.168	L	I.E.	I.E.	1
	I471T (*CYP2A6*7*)	Unknown		5%[d]	L	0.119	L	L	D	2
	G479V (*CYP2A6*5*)	Unknown		0%[d]	L	0.549 (0.172, 0.925)	G	I.E.	I.E.	1

Note: CYP2A6/POR: ratio of CYP2A6 protein expression relative to POR.

V_{max}/K_m: measure of catalytic efficiency based on in vitro metabolism to the nicotine substrate. X: Premature stop Codon. L: loss-of-function, D: decrease-of-function, N: neutral-function G: gain-of-function. Aggregate Functional Assignment (4-level): overall functional assignment of variants split into four groups: L, D, N, and G. Aggregate Functional Assignment (3-level): overall functional assignment of variants split into three groups: D, N, and G. I.E.: inconsistent evidence between in vitro and in vivo parameters. Cut points for definitions were based on reference constructs and were as follows. CYP2A6/POR: L: 0–0.4, D: 0.4–0.8, N: 0.8–1.1, G: >1.1. V_{max}/K_m: L: 0–55, D: 55–110, N:110–160, G: >160. Average NMR: L: 0–0.2, D: 0.2–0.35, N: 0.35–0.5, G: >0.5.

Abbreviations: AFR, African ancestry; EUR, European ancestry; K_m, kinetic metabolite; NMR, nicotine metabolite ratio; V_{max}, maximum value; WT, wildtype.

[a]Only individuals that were heterozygote for the allele were considered (i.e. *CYP2A6 *1/*17*, and not **17/*17*).

[b]All coding variants in the haplotype were required to be considered part of the known * allele.

[c]Reported in Mwenifumbo et al.[33]

[d]Reported in Hosono et al.[26]

[e]Reported in Piliguian et al.[24]

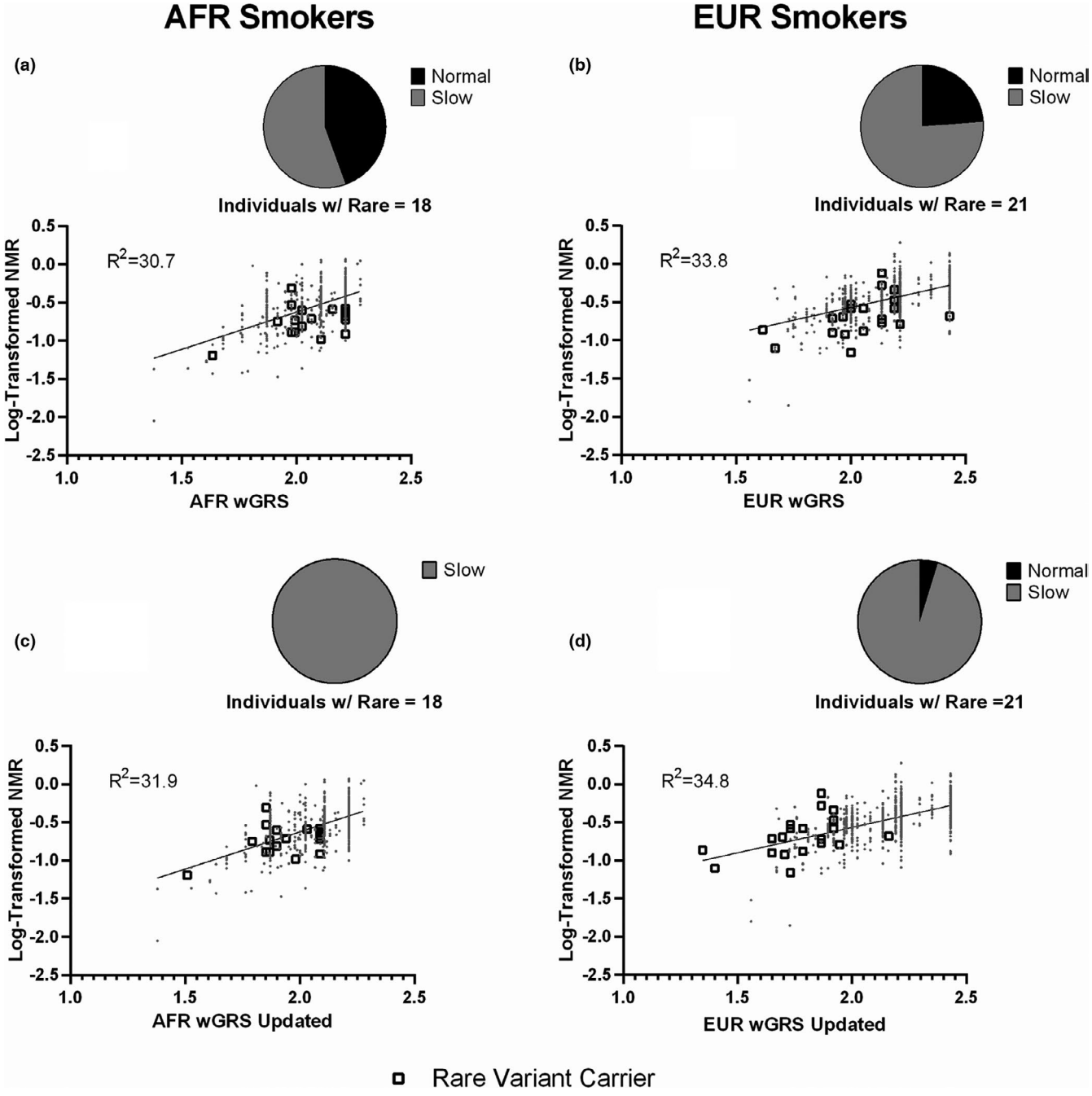
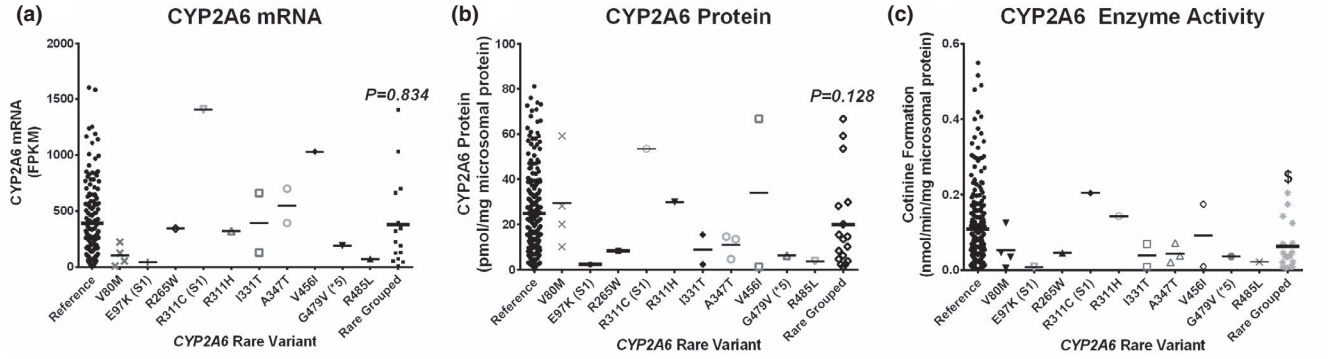


FIGURE 3 Respective (a) AFR wGRS ($N = 504$) and (b) EUR wGRS ($N = 933$), individuals with rare coding variants highlighted. Original wGRSs were based exclusively on common variants. (c, d) Updated wGRSs incorporating the functional rare coding variants weighted collectively as a decrease-of-function in the AFR and EUR population, respectively. Pie charts represent the proportion of individuals classified as slow or normal metabolizers by the wGRS before (a), (b), and after (c), (d) incorporating rare variant data. wGRS cut points for metabolizer status determination have been previously described (slow wGRS <2.089; normal wGRS ≥2.089 in AFR[18]; slow wGRS <2.14; normal wGRS ≥2.14 in EUR[17]). AFR, African-ancestry; EUR, European-ancestry; NMR, nicotine metabolite ratio; wGRS, weighted Genetic Risk Score

silico functional predictions. In silico variant predictions, even when aggregating a collection of approaches, were ineffective in distinguishing relationships with the NMR (Figure S3). The in vitro characterizations of variant effects on CYP2A6 activity were more consistent with in vivo measures of CYP2A6 activity (as demonstrated here and previously[24-26]). Third, future sequencing and analyses should explore the role of noncoding rare coding variants, such as those in regulatory elements including promotor regions which can affect the expression of *CYP2A6*. Another source of CYP2A6 functional variability may be explained by rare genetic variants in other genes



(d)	Variant[1]	mRNA (FPKM)		Protein (pmol/mg)		Enzyme Activity (nmol/min/mg)		Aggregate Functional Assignment (4-level)	Aggregate Functional Assignment (3-level)	Level of Evidence
Reference	WT	391	N	25	N	0.11	N	N	N	-
	L160H (*2)	479	N	13	L	0.06	L	L	D	-
	TATA Box (*9)	375	D	21	N	0.09	D	D	D	-
Study 1 Novel	E97K	44	L	2	L	0.01	L	L	D	3
	R311C	1405	G	53	G	0.20	G	G	G	3
Novel Rare	V80M	102	L	29	N	0.05	L	L	D	2
	R265W	344	D	8	L	0.05	L	L	D	3
	I331T	394	N	9	L	0.04	L	L	D	2
	A347T	546	N	11	L	0.04	L	L	D	2
	R485L	72	L	4	L	0.02	L	L	D	3
	V456I	1030	G	34	G	0.09	D	I.E.	I.E.	1
	R311H	320	D	30	G	0.14	N	I.E.	I.E.	1
Known Rare	G479V (*5)	189[2]	L	6[2]	L	0.04	L	L	D	3

[1] Only individuals that were heterozygote for the allele were considered (i.e. **1/*2, *1/*9* and not **2/*2* or **9/*9*)
[2] Values measured in Study Two

FIGURE 4 Association of rare *CYP2A6* coding variants with (a) CYP2A6 mRNA levels (FPKM values), (b) CYP2A6 protein levels (pmol/mg microsomal protein), and (c) CYP2A6 enzyme activity (cotinine formation from nicotine, nmol/min/mg microsomal protein) in the human liver bank ($N = 312$). Reference: Excluding those with a *CYP2A6* rare variant and/or common functional *CYP2A6* * alleles. S1: Variant also identified in Smoking Cessation Trial. Statistical analyses were based off the SKAT test statistic for continuous traits comparing the Rare Grouped and Reference groups. $: $p < 0.05$. (d) Summary of the assessments in the human liver bank. Enzyme activity, cotinine formation from nicotine in vitro; L, loss-of-function; D, decrease-of-function; N, neutral-function; G, gain-of-function. Aggregate Functional Assignment (4-level): overall functional assignment of variants split into four groups: L, D, N, and G. Aggregate Functional Assignment (3-level): overall functional assignment of variants split into three groups: D, N, and G. I.E.: inconsistent evidence between expression and kinetic parameters. Cut points for definitions were based on reference variants and were as follows. Defining parameters: mRNA: L: 0–250, D: 251–375, N: 376–600, G: greater than 600. Protein: L: 0–15, D: 16–20, N: 21–30, G: greater than 30. Enzyme Activity: L: 0–0.08, D: 0.09–0.10, N: 0.11–0.15, G: greater than 0.15

that affect the regulation or activity of CYP2A6, for example, the gene encoding POR, which is responsible for the electron transfer between NADPH and CYP enzymes.

In the Smoking Cessation Trial, a greater number of rare coding variants per capita was identified in the AFR population than in the EUR population (Table 1), as noted in other studies.[36] This reflects the greater genetic diversity found in AFR populations in general,[37] and is consistent with the on average lower NMR in AFR compared to EUR, partially explained by the greater frequency of variants in AFR individuals.[38]

An important role of rare coding variants in mediating smoking behaviors has also been shown for the genes encoding nicotine acetylcholine receptor alpha subunits (*CHRNA*). Rare coding variants identified in *CHRNA4*[39] were associated with nicotine dependence measures and smoking-related disease risk.[39] As the NMR is also associated with several smoking behaviors, including nicotine dependence measures and disease risk, we predict functionally relevant *CYP2A6* rare coding variants would yield similar results. Indeed, in a previous study involving deep sequencing the *CYP2A6* gene, there was an association between smoking amount and rare *CYP2A6* variants predicted to be deleterious based on in silico predictions[40]; these analyses could be strengthened through the integration of in vitro functional characterization, as presented in this study. Due to our focus on coding rare coding variants, there was only a small number of participants with functionally important rare coding variants in our study, thus we were underpowered to study associations with

clinical smoking phenotypes, such as smoking quantity or cessation; inclusion of functional noncoding rare coding variants would provide greater power to explore clinical associations.

For our in vitro functional characterization in the Smoking Cessation Trial, we used an *E. coli* expression system. Whereas a major advantage of this expression system is a high protein yield that can be used for structural or metabolic assays, a disadvantage are the differences in transcriptional/translational/post-translational mechanisms in bacterial versus human cells, which could influence effects on enzyme protein levels. Despite this limitation, this system has been shown here and elsewhere to reflect in vivo CYP2A6 activity.[24] The use of RNA-seq to confirm rare coding variants identified in the Human Liver Bank is not as reliable as Sanger sequencing, as was done to confirm variants in the Smoking Cessation Trial; however, unique primer sets were used between the DNA and RNA sequencing, making it less likely for the two orthogonal approaches to result in identical variant calling errors. Due to the cDNA nature of our in vitro functional assay, we focused our assessments on coding variants; however, it is possible that rare coding variants within regulatory elements and splice sites could contribute to the functional variability in CYP2A6 and the NMR. Future studies should consider the implications of these noncoding rare coding variants. Finally, for our in vitro assessments all variants were characterized in individual constructs, however, as displayed in Table S2, it is possible that variant combinations could lead to alternative effects on CYP2A6 activity that were not explicitly tested for. However, the rare variants were most often found in the same individuals as *CYP2A6*9*, a variant in the 5′ regulatory element (TATA box) that affects expression which could not be explicitly tested in our *E. coli* (prokaryote) expression system. Furthermore, the integration of the genetic risk scores in Figure 3 accounts for circumstances where individuals also have other functional variants.

Considering there are 494 amino acids in the CYP2A6 protein with thousands of potential coding variant possibilities, expanding the catalogue of functionally characterized variants will be necessary. One approach would be through the techniques presented here; another approach would be to implement a deep mutational scanning approach. Yeast-based activity assays to test thousands of variants are being piloted in *CYP2C9*, although the reflection of these assays to gold standard in vitro metabolic assays is still undetermined.[41] Alternatively, current in silico prediction techniques could undergo further refinement with the growing knowledge of variant effects.

In conclusion, we demonstrate that incorporation of functionally relevant rare coding variants can be important for the determination of individual metabolizer status. Incorporation of these rare coding variants led to more accurate CYP2A6 activity grouping (normal vs. slow) for individuals genotyped with rare coding variants. Improved CYP2A6 groupings can in turn provide better assessments of the impact of CYP2A6 on outcomes of interest, including smoking behaviors, disease risk assessment, as well as improving tailored selection of smoking cessation treatment. Currently, the high costs and poor accuracy of sequencing may prevent integration of rare coding variants, but as the accuracy, frequency, and appeal of sequencing increases, to achieve precision medicine for all, integration of these variants should be considered.

ACKNOWLEDGEMENTS

Computations were performed on the Centre for Addiction and Mental Health (CAMH) Specialized Computing Cluster (SCC), funded by the Canada Foundation for Innovation Research Hospital Fund. We acknowledge additional members of the PGRN-PNAT Research Group including Frank Leone, Henry Glick, Angela Pinto, Paul Sanborn, Peter Gariti, Richard Landis (University of Pennsylvania); Maria Novalen, Bin Zhao, Ewa Hoffmann, Qian Zhou, Adel Aziziyeh (CAMH/University of Toronto); Martin Mahoney (Roswell Cancer Center, University of Buffalo); Maher Karam-Hage (The University of Texas M.D. Anderson Cancer Center); David Conti (University of Southern California); and Andrew Bergen (SRI International). We would like to thank Jennie Pouget for proofreading the manuscript. This publication was made possible by the Pharmacogenomics Research Network-RIKEN Global Alliance (PGRN-RIKEN), which is supported by the RIKEN Center for Integrative Medical Science and the NIH Pharmacogenomics Research Network (GM115370).

CONFLICTS OF INTEREST

R.F.T. has consulted for Quinn Emanuel and Ethismos on unrelated topics. N.L.B. has been a consultant to Pfizer and Achieve Life Sciences, companies that market or are developing smoking cessation medications, and has been an expert witness in litigation against tobacco companies. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

A.E. and R.F.T. wrote the manuscript. A.E., A.Z.X.Z., K.F., T.M., M.K., and R.F.T. designed the research. A.E., J.T., K.G.C., B.P., and K.F. performed the research. A.E. analyzed the data. E.G.S., K.E.T., T.M., M.K., N.L.B., C.L., and R.F.T. contributed reagents/analytical tools.

ORCID

Katrina G. Claw https://orcid.org/0000-0003-2239-5018
Rachel F. Tyndale https://orcid.org/0000-0003-1297-2053

REFERENCES

1. Benowitz NL. Nicotine addiction. *N Engl J Med*. 2010;362:2295-2303.
2. Nakajima M, Yamamoto T, Nunoya K, et al. Role of human cytochrome P4502A6 in C-oxidation of nicotine. *Drug Metab Dispos*. 1996;24:1212-1217.
3. Dempsey D, Tutka P, Jacob P 3rd,et al. Nicotine metabolite ratio as an index of cytochrome P450 2A6 metabolic activity. *Clin Pharmacol Ther*. 2004;76:64-72.
4. O'Loughlin J, Paradis G, Kim W, et al. Genetically decreased CYP2A6 and the risk of tobacco dependence: a prospective study of novice smokers. *Tob Control*. 2004;13:422-428.
5. Schoedel KA, Hoffmann EB, Rao Y, Sellers EM, Tyndale RF. Ethnic variation in CYP2A6 and association of genetically slow nicotine metabolism and smoking in adult Caucasians. *Pharmacogenetics*. 2004;14:615-626.
6. Wassenaar CA, Dong Q, Wei Q, et al. Relationship between CYP2A6 and CHRNA5-CHRNA3-CHRNB4 variation and smoking behaviors and lung cancer risk. *J Natl Cancer Inst*. 2011;103:1342-1346.
7. Strasser AA, Benowitz NL, Pinto AG, et al. Nicotine metabolite ratio predicts smoking topography and carcinogen biomarker level. *Cancer Epidemiol Biomarkers Prev*. 2011;20:234-238.
8. Wassenaar CA, Ye Y, Cai Q, et al. CYP2A6 reduced activity gene variants confer reduction in lung cancer risk in African American smokers–findings from two independent populations. *Carcinogenesis*. 2015;36:99-103.
9. Patterson F, Schnoll RA, Wileyto EP, et al. Toward personalized therapy for smoking cessation: a randomized placebo-controlled trial of bupropion. *Clin Pharmacol Ther*. 2008;84:320-325.
10. Lerman C, Schnoll RA, Hawk LW, et al. Use of the nicotine metabolite ratio as a genetically informed biomarker of response to nicotine patch or varenicline for smoking cessation: a randomised, double-blind placebo-controlled trial. *Lancet Respir Med*. 2015;3:131-138.
11. Lerman C, Jepson C, Wileyto EP, et al. Genetic variation in nicotine metabolism predicts the efficacy of extended-duration transdermal nicotine therapy. *Clin Pharmacol Ther*. 2010;87:553-557.
12. Lerman C, Tyndale R, Patterson F, et al. Nicotine metabolite ratio predicts efficacy of transdermal nicotine for smoking cessation. *Clin Pharmacol Ther*. 2006;79:600-608.
13. McDonagh EM, Wassenaar C, David SP, et al. PharmGKB summary: very important pharmacogene information for cytochrome P-450, family 2, subfamily A, polypeptide 6. *Pharmacogenet Genomics*. 2012;22:695-708.
14. Swan GE, Lessov-Schlaggar CN, Bergen AW, He Y, Tyndale RF, Benowitz NL. Genetic and environmental influences on the ratio of 3′hydroxycotinine to cotinine in plasma and urine. *Pharmacogenet Genomics*. 2009;19:388-398.
15. Loukola A, Buchwald J, Gupta R, et al. A genome-wide association study of a biomarker of nicotine metabolism. *PLoS Genet*. 2015;11:e1005498.
16. Buchwald J, Chenoweth MJ, Palviainen T, et al. Genome-wide association meta-analysis of nicotine metabolism and cigarette consumption measures in smokers of European descent [published online ahead of print March 10, 2020]. *Mol Psychiatry*. 2020. 10.1038/s41380-020-0702-z
17. El-Boraie A, Taghavi T, Chenoweth MJ, et al. Evaluation of a weighted genetic risk score for the prediction of biomarkers of CYP2A6 activity. *Addict Biol*. 2020;25:e12741.
18. El-Boraie A, Chenoweth MJ, Pouget JG, et al. Transferability of ancestry-specific and cross-ancestry CYP2A6 activity genetic risk scores in African and European populations [published online ahead of print December 9, 2020]. *Clin Pharmacol Ther*. 2021. 10.1002/cpt.2135
19. Silventoinen K, Sammalisto S, Perola M, et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res*. 2003;6:399-408.
20. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747-753.
21. Kozyra M, Ingelman-Sundberg M, Lauschke VM. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet Med*. 2017;19:20-29.
22. Raunio H, Rautio A, Gullstén H, Pelkonen O. Polymorphisms of CYP2A6 and its practical consequences. *Br J Clin Pharmacol*. 2001;52:357-363.
23. Zhou Y, Mkrtchian S, Kumondai M, Hiratsuka M, Lauschke VM. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J*. 2019;19:115-126.
24. Piliguian M, Zhu AZX, Zhou Q, et al. Novel CYP2A6 variants identified in African Americans are associated with slow nicotine metabolism in vitro and in vivo. *Pharmacogenet Genomics*. 2014;24:118-128.
25. Ho MK, Mwenifumbo JC, Zhao B, Gillam EM, Tyndale RF. A novel CYP2A6 allele, CYP2A6*23, impairs enzyme function in vitro and in vivo and decreases smoking in a population of Black-African descent. *Pharmacogenet Genomics*. 2008;18:67-75.
26. Hosono H, Kumondai M, Maekawa M, et al. Functional characterization of 34 CYP2A6 allelic variants by assessment of nicotine C-oxidation and coumarin 7-hydroxylation activities. *Drug Metab Dispos*. 2017;45:279-285.
27. Tanner JA, Prasad B, Claw KG, et al. Predictors of variation in CYP2A6 mRNA, protein, and enzyme activity in a human liver bank: influence of genetic and nongenetic factors. *J Pharmacol Exp Ther*. 2017;360:129-139.
28. Twist GP, Gaedigk A, Miller NA, et al. Constellation: a tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene, CYP2D6, from whole-genome sequences. *NPJ Genom Med*. 2016;1:15007.
29. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
30. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297-1303.
31. Mu W, Lu HM, Chen J, Li S, Elliott AM. Sanger confirmation is required to achieve optimal sensitivity and specificity in next-generation sequencing panel testing. *J Mol Diagn*. 2016;18:923-932.
32. St. Helen G, Novalen M, Heitjan DF, et al. Reproducibility of the nicotine metabolite ratio in cigarette smokers. *Cancer Epidemiol Biomarkers Prev*. 2012;21:1105-1114.

33. Mwenifumbo JC, Al Koudsi N, Ho MK, et al. Novel and established CYP2A6 alleles impair in vivo nicotine metabolism in a population of Black African descent. *Hum Mutat*. 2008;29:679-688.
34. Tanner JA, Zhu AZ, Claw KG, et al. Novel CYP2A6 diplotypes identified through next-generation sequencing are associated with in-vitro and in-vivo nicotine metabolism. *Pharmacogenet Genomics*. 2018;28:7-16.
35. Basile AO, Byrska-Bishop M, Wallace J, Frase AT, Ritchie MD. Novel features and enhancements in BioBin, a tool for the biologically inspired binning and association analysis of rare variants. *Bioinformatics*. 2018;34:527-529.
36. Gordon AS, Tabor HK, Johnson AD, et al. Quantifying rare, deleterious variation in 12 human cytochrome P450 drug-metabolism genes in a large-scale exome dataset. *Hum Mol Genet*. 2014;23:1957-1963.
37. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet*. 2008;9:403-433.
38. Ross KC, Gubner NR, Tyndale RF, et al. Racial differences in the relationship between rate of nicotine metabolism and nicotine intake from cigarette smoking. *Pharmacol Biochem Behav*. 2016;148:1-7.
39. Thorgeirsson TE, Steinberg S, Reginsson GW, et al. A rare missense mutation in CHRNA4 associates with smoking behavior and its consequences. *Mol Psychiatry*. 2016;21:594-600.
40. Clark SL, McClay JL, Adkins DE, et al. Deep sequencing of three loci implicated in large-scale genome-wide association study smoking meta-analyses. *Nicotine Tob Res*. 2016;18:626-631.
41. Amorosi C, Chiasson MA, McDonald MG, et al. Massively parallel functional profiling of CYP2C9 variants using a yeast activity assay. (Program Number: 318). *Annual Meeting of the American Society of Human Genetics*. 2019;108(96):P1735-P1751.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: El-Boraie A, Tanner J-A, Zhu AZX, et al. Functional characterization of novel rare *CYP2A6* variants and potential implications for clinical outcomes. *Clin Transl Sci*. 2022;15:204–220. https://doi.org/10.1111/cts.13135