10-2024

PharmVar GeneFocus: CYP2A6.

Alec W R Langlois

Meghan J. Chenoweth

David Twesigomwe

Giada Scantamburlo

Michelle Whirl-Carrillo

See next page for additional authors

Recommended Citation

Langlois AWR, Chenoweth MJ, Twesigomwe D, et al. PharmVar GeneFocus: CYP2A6. Clin Pharmacol Ther. 2024;116(4):948-962. doi:10.1002/cpt.3387

PharmVar GeneFocus: *CYP2A6*

Alec W.R. Langlois[1,2], Meghan J. Chenoweth[1,2,3], David Twesigomwe[4], Giada Scantamburlo[5], Michelle Whirl-Carrillo[6], Katrin Sangkuhl[6], Teri E. Klein[6,7], Charity Nofziger[5], Rachel F. Tyndale[1,2,3] and Andrea Gaedigk[8,9,*]

The Pharmacogene Variation Consortium (PharmVar) provides nomenclature for the human *CYP2A* gene locus containing the highly polymorphic *CYP2A6* gene. CYP2A6 plays a role in the metabolism of nicotine and various drugs. Thus, genetic variation can substantially contribute to the function of this enzyme and associated efficacy and safety. This GeneFocus provides an overview of the clinical significance of *CYP2A6*, including its genetic variation and function. We also highlight and discuss caveats in the identification and characterization of allelic variation of this complex pharmacogene, a prerequisite for accurate genotype determination and prediction of phenotype status.

INTRODUCTION, BRIEF HISTORY, AND STATUS OF NOMENCLATURE BEFORE PHARMVAR

CYP2A6, encoded by the *CYP2A6* gene, is the primary enzyme involved in the metabolic inactivation of nicotine in humans. CYP2A6's discovery was rooted in early research on nicotine metabolism. By the late 1950s, work in dogs and human smokers identified cotinine and hydroxycotinine as metabolites of nicotine, and suggested the conversion of nicotine to cotinine and further to hydroxycotinine as a potential pathway for the elimination of nicotine *in vivo*.[1] In the 1970s, the oxidation of nicotine to an iminium ion was discovered as an intermediate step before conversion to cotinine, and it was proposed that the enzyme catalyzing this intermediate step was a mixed-function oxidase.[1]

During the 1980s, many cytochrome P450 (CYP) proteins were purified from rat livers and assayed for substrate specificity. Antibodies targeting a rat CYP with high steroid hydroxylation activity (initially named P-450a, now known as *cyp2A1*) were used to isolate cDNA for a human CYP isoform from a liver cDNA expression library.[2] The CYP, initially named IIA3 or CYP2A3, and later renamed CYP2A6, was the primary coumarin 7-hydroxylating enzyme in the human liver.[3] Early investigations of CYP2A6 often used coumarin as a probe substrate.[3,4]

By the 1990s, the crucial role of CYP2A6 in the metabolism of nicotine to an iminium ion, which is then converted to cotinine by cytosolic aldehyde oxidase, then sequentially to 3′-hydroxycotinine (3'-HC), was discovered; of note CYP2A6 can form cotinine from nicotine in the absence of cytosolic aldehyde oxidase, as demonstrated *in vitro* using purified CYP2A6 expressed from cDNA in *E. coli* without added cytosol.[1,5] Further, the conversion of nicotine to cotinine was found to be the major pathway of nicotine metabolic clearance *in vivo*.[1] Around the same time, CYP2A6 was implicated in the activation of pre-carcinogenic nicotine-related nitrosamines found in cigarette smoke.[6]

In 1995, the *CYP2A6* gene was identified and localized to a cluster of *CYP2A*, *CYP2B*, *CYP2F*, and other *CYP* genes on chromosome 19, enabling sequencing and the discovery of variant alleles.[4] High interindividual variability in the rate of cotinine formation from nicotine in human liver microsomes was consistent with the existence of *CYP2A6* variant alleles.[7,8] The nearby homologues *CYP2A7* (inactive) and *CYP2A13* (active, described in the "Brief Overview of CYP2A13" section below), and their high nucleotide sequence identity with *CYP2A6* (all three mRNAs > 90% identical with each other), were also identified at this time.[4]

Two *CYP2A6* variant alleles were characterized in the mid-1990s and initially named *CYP2A6v1* and *CYP2A6v2*.[4] Shortly thereafter, the star allele nomenclature system was adopted for *CYP2A6*, and *CYP2A6v1* and *CYP2A6v2* were renamed *CYP2A6*2* and *CYP2A6*3*. While *CYP2A6*2* remains an important functional star allele today, *CYP2A6*3* was soon discovered to be a sequencing artifact arising due to the placement of primers within a common *CYP2A6* 3′UTR conversion to *CYP2A7*.[9] Initial analyses of the impact of *CYP2A6* variant alleles on smoking behaviors in the late 1990s established the strong association between inactive *CYP2A6* alleles (such as *CYP2A6*2*) and decreased cigarettes smoked per day, as well as decreased risk of smoking dependence.[10]

In 2000, the Human Cytochrome P450 Allele Nomenclature website was launched, and eventually two more star alleles, *CYP2A6*4* (a gene deletion, initially called *CYP2A6del*) and *CYP2A6*5* (characterized by an amino acid changing variant in exon 9)[9] were posted. By the time *CYP2A6* was curated by the expert panel and transferred into the PharmVar database in 2022, 45

[1]Department of Pharmacology & Toxicology, University of Toronto, Toronto, Ontario, Canada; [2]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; [3]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; [4]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; [5]PharmGenetix GmbH, Niederalm, Austria; [6]Department of Biomedical Data Science, Stanford University, Stanford, California, USA; [7]Departments of Medicine (BMIR) and Genetics, Stanford University, Stanford, California, USA; [8]Division of Clinical Pharmacology, Toxicology and Therapeutic Innovation, Children's Mercy Research Institute (CMRI), Kansas City, Missouri, USA; [9]School of Medicine, University of Missouri-Kansas City, Kansas City, Missouri, USA. *Correspondence: Andrea Gaedigk (agaedigk@cmh.edu)

Received May 1, 2024; accepted July 8, 2024. doi:10.1002/cpt.3387

unique star alleles (and many more suballeles) had been identified. Star alleles are mostly defined by single nucleotide polymorphisms and/or small insertions and deletions which are henceforth collectively referred to as single nucleotide variants (SNVs) or simply variants.

CLINICAL RELEVANCE

CYP2A6 metabolizes a variety of drugs, including efavirenz, tegafur, letrozole, coumarin, and metronidazole.[6] CYP2A6 also metabolizes several nicotine-derived pre-carcinogenic nitrosamines including N′-nitrosonornicotine (NNN) and 4-(methylnitrosamino-1-(3-pyridyl)-butanone) (NNK); expressed *CYP2A6* metabolizes NNN more efficiently than NNK.[11,12] However, nicotine and cotinine remain the most studied CYP2A6 substrates due to nicotine's relationship with cigarette smoking and subsequent tobacco-related diseases, as well as cotinine's importance as a biomarker of nicotine exposure and the denominator in the nicotine metabolite ratio (NMR; see below).

Nicotine has a relatively short half-life (2 hours) resulting in the need to smoke at regular short intervals.[13] Cotinine has a longer half-life (16 hours)[13] compared to nicotine and 3′-HC (6 hours),[1] and CYP2A6 is exclusively responsible for the formation-dependent kinetics of 3′-HC.[14,15] Thus, CYP2A6 activity can be assayed in *ad libitum* smokers by calculating the ratio of the concentrations of 3′-HC to cotinine in plasma (or blood, saliva or urine); this ratio is commonly known as the nicotine metabolite ratio (NMR).[15,16] The NMR is highly correlated with nicotine clearance ($r = 0.90$), and is related to *CYP2A6* genotype due to CYP2A6's major role in nicotine metabolism.[15] While the NMR is most highly correlated with nicotine clearance *in vivo* when measured in plasma (or blood or saliva), the urinary NMR is only marginally less correlated when measured as $3'\text{-HC}_{total}/\text{cotinine}_{free}$, where $3'\text{-HC}_{total}$ includes both free and glucuronidated 3′-HC; this is consistent with the fact that CYP2A6 metabolizes only free cotinine (to 3′-HC, which can then be glucuronidated or free).[16] However, the correlation with nicotine clearance drops when using urinary $3'\text{-HC}_{total}/\text{cotinine}_{total}$ (as cotinine glucuronide is not a CYP2A6 substrate) or $3'\text{-HC}_{free}/\text{cotinine}_{free}$ (as this does not capture the sum of CYP2A6 downstream metabolites).[16]

In regular smokers, the NMR is robust (e.g., to time of day sampling and to smoking level) and stable (e.g., at room temperature over many days), allowing for assay of CYP2A6 activity with a single measurement.[15,17] Lower NMR (i.e., slower CYP2A6 activity) and/or reduced activity *CYP2A6* genotypes are associated with many smoking behaviors, including smoking fewer cigarettes per day, lower nicotine reinforcement, lower risk of dependence, lower puff volume, and greater rates of smoking cessation.[18] As established in many ancestral groups, lower NMR and/or reduced activity *CYP2A6* genotypes are also associated with increased BMI,[19] as is a lower risk of COPD and lung cancer,[20–22] primarily through the impact on smoking and possibly, for tobacco-related cancers, through CYP2A6's role in nitrosamine activation. Since CYP2A6 and the NMR do not have standardized phenotype categories (see "Genotype to Phenotype Translation" section), this review uses "faster" or "slower" to describe NMR groupings when activity was grouped, or to indicate directionality when the NMR was used continuously.

Nicotine, a full nicotinic receptor agonist, is itself used as a smoking cessation therapeutic known as nicotine replacement therapy (NRT), along with the nicotinic receptor antagonist bupropion and partial agonist varenicline. CYP2A6 activity is associated with the relative efficacy of these therapeutics. Most clinical trials (all trials include behavioral counseling) have shown that slower metabolizers have better smoking cessation rates than faster metabolizers when treated with a placebo or nicotine patch, even when extending the duration of NRT treatment.[23–26] While quit rates were similar on the nicotine nasal spray, a titratable nicotine treatment, slower metabolizers used less nicotine than faster metabolizers, similar to the titration observed when acquiring nicotine via smoking.[27] In a prospectively NMR-stratified cessation trial, faster metabolizers were more likely to quit successfully using varenicline than nicotine patch, while slower metabolizers had similar quit rates but experienced more side effects on varenicline than on nicotine patch.[23] This led to the suggestion that faster metabolizers should be preferentially treated with varenicline while slower metabolizers should be treated with nicotine patch.[23] Similar quit rates were found for slower and faster nicotine metabolizers on bupropion (which is a CYP2B6 substrate and not a CYP2A6 substrate).[26]

CYP2A6 is also relevant in the co-use of alcohol and nicotine. Heavy alcohol use and smoking are highly co-morbid, and heavy drinkers are twice as likely to smoke vs. non-heavy drinkers.[28] A study of alcohol cessation found that CYP2A6 activity (measured using the NMR) decreased substantially by 7 weeks after alcohol cessation, implying that chronic alcohol use causes higher CYP2A6 activity.[29] Given that faster CYP2A6 activity is associated with lower rates of smoking cessation (see above) and that chronic alcohol users also have lower rates of smoking cessation, it is proposed that faster CYP2A6 activity among heavy drinkers contributes to difficulties in quitting smoking.[29]

CYP2A6 has been implicated both in HIV pathogenesis and higher smoking among HIV-positive individuals (vs. HIV-negative individuals).[30] Given the use of modern antiretroviral treatments, HIV-positive individuals are now more likely to die as a result of smoking as opposed to AIDS.[31] Smokers with HIV have higher NMRs than non-HIV smokers, as well as higher NMRs after viral suppression using antiretrovirals (vs. before viral suppression).[32,33]

HIV infection itself does not explain the higher NMRs.[34] Rather, the use of the antiretroviral HIV treatment efavirenz (a CYP2A6 substrate) likely causes the increase in CYP2A6 activity, since switching from efavirenz to a different antiretroviral regimen decreased NMRs.[32,35] Consistent with these findings, a study using a caffeine metabolite as a probe for CYP2A6 showed that 17 days of efavirenz treatment nearly doubled CYP2A6 activity, an effect that may occur via activation of the nuclear pregnane X (PXR; *NR1I2*) and constitutive androstane receptors (CAR; *NR1I3*) and subsequent CYP2A6 induction.[36] Significant elevation of NMRs is also seen in patients on other antiretroviral regimens, although increases are more minor compared to those on efavirenz; decreased inflammation following viral suppression may explain these findings.[32,33] Overall, faster CYP2A6 activity may

reduce smoking cessation among HIV-positive individuals, thereby contributing to the greater prevalence of smoking.

CYP2A6 is also involved in the metabolism of chemotherapeutics such as tegafur and letrozole. Genetic variation in *CYP2A6* is associated with variable pharmacokinetics of tegafur, as this drug is metabolized by CYP2A6 into an active metabolite, 5′-fluorouracil (5-FU), which is also a CYP2A6 substrate.[6] Several studies have found higher response rates to S-1 (a gastric cancer combination pharmacotherapy that includes tegafur) in gastric cancer patients with faster CYP2A6 metabolism compared to those with *CYP2A6* genotypes imparting slower metabolism.[37,38]

While genetic variation in *CYP2A6* has been shown to impact *in vivo* drug levels for a variety of other CYP2A6 substrates (efavirenz, letrozole, valproic acid, dexmedetomidine, etc.), there is currently limited evidence to establish a clinically significant impact on these substrates.[6]

AGE-RELATED AND SEX DIFFERENCES

CYP2A6 is expressed during fetal development, and most publications, whether *in vivo* or *in vitro*, find little or no evidence of changes in CYP2A6 expression or activity from infancy through childhood, adulthood, and into old age (except as a result of sex differences in females, see below).[6,8,39] Of note, increased nicotine half-life has been observed in infants *in vivo* (vs. adults), but with no difference in cotinine half-life.[40] Nicotine has high hepatic extraction relative to cotinine suggesting that differences in nicotine half-life in infants are likely due to differences in hepatic blood flow rather than developmental changes in CYP2A6 activity.[13] Likewise, the slower nicotine clearance observed in the elderly is likely due to decreased hepatic blood flow rather than lower CYP2A6 activity.[41]

As mentioned briefly above, CYP2A6 expression is induced by estrogen through an interaction with a proximal (~2.4 kb upstream) estrogen response element.[42] In human liver banks, adult female livers have higher mRNA and protein expression vs. adult male livers.[39,43] This results in faster CYP2A6 activity in females (vs. males), pre-menopausal females (vs. post-menopausal females), and among users of oral contraceptives or hormone replacement therapies containing estrogen (vs. those not containing estrogen), while there is no difference in CYP2A6 activity between males and post-menopausal females.[19,44,45]

The pharmacokinetics of CYP2A6 substrates in infants and children exposed to second-hand smoke in the household have also been investigated, with no sex differences in the half-life of cotinine or NMRs being observed.[46] This is consistent with the hypothesis that sex differences in CYP2A6 activity appear as estrogen increases among females during adolescence and disappear when estrogen decreases post-menopause.

Urinary NMRs ($3'\text{-HC}_{total}/\text{cotinine}_{free}$) showed an increase in CYP2A6 activity in early pregnancy, which is even greater in late pregnancy, relative to 6 months postpartum,[47,48] consistent with estrogen induction of CYP2A6 and the known increase in estrogen concentrations over the course of pregnancy. Various nicotine metabolic pathways were investigated during pregnancy using appropriate metabolite ratios, and the activity of pathways involving CYP2A6, as well as nicotine and cotinine glucuronidation, increased across the course of pregnancy.[47–49] Administration of progesterone to human hepatocytes also resulted in the induction of CYP2A6, albeit to a lesser extent than estrogen, implying that progesterone also contributes to the activity changes observed during pregnancy.[50] In pregnancy, there is little evidence for the benefit of NRT.[51] Rather, it is suggested that an increase in CYP2A6 activity during pregnancy may reduce the therapeutic levels of the NRT.[48]

OTHER FACTORS INFLUENCING CYP2A6 ACTIVITY

While most of the variation in CYP2A6 activity is heritable (60–80% using the NMR),[6] there are several other sources of variation, including both drug and hormonal inducers. CYP2A6 is induced by estrogen (described above in "Age-related and sex differences"), phenobarbital, pyrazole, and rifampin in cultured human hepatocytes, and phenobarbital and dexamethasone in human liver microsomes.[6] However, in one liver bank study, livers exposed to phenobarbital or dexamethasone did not show significantly higher CYP2A6 expression or activity.[39] These induction effects are thought to involve the activation of nuclear receptors, including the PXR (through a 5′ DR4-like response element), the CAR, and the glucocorticoid receptor (GR; NR3C1).[6] As already mentioned above, efavirenz, along with other antiretrovirals albeit to a lesser extent, is associated with a substantial increase in CYP2A6 activity.[32]

CYP2A6 is also susceptible to drug inhibition. Methoxsalen, selegiline, and isoniazid are all mechanism-based inhibitors (as shown in human liver microsomes).[6] The inhibitory activity of methoxsalen has also been observed *in vivo*, where its administration in smokers inhibited nicotine metabolism (i.e., increased nicotine concentration) and reduced smoking.[52] These studies support a causal relationship between CYP2A6 activity and smoking.

In vitro experiments identified several competitive inhibitors of CYP2A6, including tranylcypromine, ketoconazole, and tryptamine, but these effects have not been measured *in vivo*.[6] CYP2A6 is inhibited *in vivo* by grapefruit juice, as well as menthol (including at concentrations found in menthol cigarettes).[6]

Cigarette smoking itself reduces nicotine metabolism, where nicotine clearance is lower while smoking[53] and higher after 1 week following cessation.[54]

NEED FOR STANDARDIZED GENETIC REPORTING AND FUNCTIONAL OR CLINICAL IMPACT

To effectively develop and implement clinical guidelines relating to *CYP2A6* variation, standardization of genetic definitions is necessary. Recent large-scale analyses have generally been carried out using array- or PCR-based genotyping of specific SNVs known or suspected to have a functional impact.[55–58] Such studies have been fruitful in explaining substantial portions of variation in CYP2A6 activity and producing tools for phenotype prediction. However, large-scale studies with both phenotype measures and full sequencing data are necessary to further elucidate the functional effects of rare and relatively low-impact SNVs.

Most analyses of CYP2A6 activity utilize the NMR. While it is a highly effective tool for the investigation of CYP2A6-mediated nicotine inactivation *in vivo*, the functional impact of *CYP2A6*

SNVs on the metabolism of other CYP2A6 substrates has, for many alleles, not been demonstrated. This is especially important given that several *CYP2A6* alleles have substrate-specific effects. For example, both *CYP2A6*7* (tested *in vivo*) and *CYP2A6*17* (tested *in vitro*) decrease nicotine metabolism but have no effect on coumarin metabolism.[59,60] Furthermore, the NMR can only be assayed in regular smokers, limiting precision medicine approaches for other CYP2A6 substrates (e.g., efavirenz, letrozole, valproic acid, dexmedetomidine) in former-, intermittent-, or non-smokers, as well as its utility for variant characterization in these groups. On another note, metronidazole has been proposed as a potentially safe and specific probe drug for use in non-smokers.[61]

Finally, many *CYP2A6* variant alleles (star alleles) are defined by multiple SNVs leading to multiple amino acid changes or altered expression levels; additionally, structural variants including gene deletions, duplications and hybrids with the *CYP2A7* pseudogene can contribute to altered activity. Thus, standardized use of star allele nomenclature, as documented on the PharmVar website and described below, is encouraged to further research and facilitate future clinical implementation.

CYP2A6 IN PHARMGKB

The PharmGKB *CYP2A6* gene page allows structured access to curated gene-specific pharmacogenomic (PGx) knowledge, including drug label annotations, clinical annotations, variant annotations, and curated pathways.[62] PharmGKB extracts relevant PGx information on approved drug labels from multiple sources, including, but not limited to, the US Food and Drug Administration (FDA) and the Pharmaceuticals and Medical Devices Agency, Japan (PMDA). PharmGKB assigns a "PGx level" to these annotations based on the amount of actionable information stated or implied on the label.

Current PharmGKB information related to CYP2A6 is summarized as follows: (i) one *CYP2A6*-related label annotation from the PMDA for letrozole with an "actionable" level due to reported pharmacokinetic differences observed in individuals with slower metabolism; (ii) 45 *CYP2A6*-related clinical annotations (CAs), and (iii) 18 pharmacokinetic pathways depicting the role of CYP2A6 in drug metabolism. CAs are evidence-rated, genotype-level summaries for specific variant or allele drug combinations based on curated literature. The CA summarizing the association between *CYP2A6* star alleles and nicotine metabolism has a high level of evidence (level 1B). Other CAs exist for non-replicated associations between *CYP2A6* SNVs and nicotine, and associations with low levels of evidence for coumarin, efavirenz, letrozole, methoxsalen, metronidazole, and tegafur. CYP2A6 plays a major role in the nicotine pharmacokinetics pathway found on PharmGKB.

THE *CYP2A6* GENE LOCUS

CYP2A6 lies in a cluster of *CYP* genes on the long arm of chromosome 19, including *CYP2A7*, *CYP2A13*, as well as *CYP2B* and *CYP2F* genes. The *CYP2A6* gene is 7 kb and includes nine exons coding for a 494 amino acid enzyme product. The gene is proximal to two homologues: *CYP2A7*, a gene encoding an inactive product 25 kb downstream of *CYP2A6*, and *CYP2A13*, a gene encoding an active enzyme 238 kb downstream of *CYP2A6* (**Figure 1**). While *CYP2A6* and *CYP2A7* are transcribed from the minus strand (i.e., transcription proceeds 5′ to 3′ with *decreasing* chromosome 19 positions), *CYP2A13* is transcribed from the plus strand (i.e., transcription proceeds 5′ to 3′ with *increasing* chromosome 19 positions).

In addition to the three intact *CYP2A* genes, a split *CYP2A* pseudogene consisting of two fragments exists downstream of *CYP2A7*. The two fragments are separated by a 114 kb region which includes *CYP2B6* and *CYP2B7*. The first fragment, named *CYP2A7P1*, has five *CYP2A*-like exons; the second fragment, named *CYP2A7P2*, has four. Together, the nine exons of *CYP2A7P1/CYP2A7P2* share 91% sequence identity with those from *CYP2A6* and *CYP2A7*.[63] Both fragments are arranged in the same orientation as *CYP2A6* and *CYP2A7* on the minus strand.

Structural variation (SV) at the *CYP2A6* gene locus is common, making up ~20% of alleles in certain populations,[6,64] and is thought to arise from unequal crossing over in regions of high sequence identity between *CYP2A6* and *CYP2A7* (introns 2, 4, and the 3′ UTR and flanking region), producing hybrid genes (*CYP2A6*12*, **34*, **53*), gene deletions (*CYP2A6*4*), and gene duplications (*CYP2A6*1x2*). Given the various similar *CYP* genes existing in the region surrounding *CYP2A6*, it is likely that *CYP2A6* itself, along with the *CYP2B*, *CYP2F*, and other *CYP2A* genes were initially generated by several gene duplication events.[63]

SNVs in *CYP2A6* exons, introns, and the 5′ and 3′UTRs and flanking regions are listed on PharmVar. Most SNVs impacting function cause amino acid changes, for example, 1799T>A (rs1801272) causing p.L160H in *CYP2A6*2*. However, as CYP2A6 expression is affected by transcriptional regulation, non-coding variants have also been shown to alter expression levels. First, −48T>G (rs28399433, *CYP2A6*9*) is an SNV in the TATA box that has been associated with decreased mRNA and protein expression.[65,66] Second, a 58-bp long conversion to *CYP2A7* sequence 3′ of the stop codon (defined by two separate rsIDs, rs386809285, rs386809286) that is found on *CYP2A6*46* and several other haplotypes causes increased expression *in vitro* and in mouse models[67]; this region is also known, and annotated by PharmVar, as the "3′UTR conversion". The 3′UTR is thought to stabilize mRNA, with *in vitro* evidence of interactions with the RNA binding protein HNRNPA1.[68] As a result of these variants, and the potential existence of other functional non-coding SNVs, PharmVar requires that all SNVs between positions −1013 and 6835 of the *CYP2A6* genomic reference sequence (RefSeq) NG_008377.1 (where ATG = +1) be included in submissions.

BRIEF OVERVIEW OF CYP2A13

CYP2A13, which has its own PharmVar page, is a relatively understudied CYP closely related to CYP2A6 (80.1% overall nucleotide identity; 94% mRNA identity; 93.5% amino acid identity). The *CYP2A13* gene produces an active enzyme with similar substrate specificity to CYP2A6; however, CYP2A13 is unlikely to contribute substantially to the systemic concentrations of substrates and metabolites (e.g., nicotine and cotinine) *in vivo*. While CYP2A6 is most highly expressed in the liver, CYP2A13 expression is lower overall and occurs primarily in the respiratory tract, where CYP2A6 is also expressed.[69] A

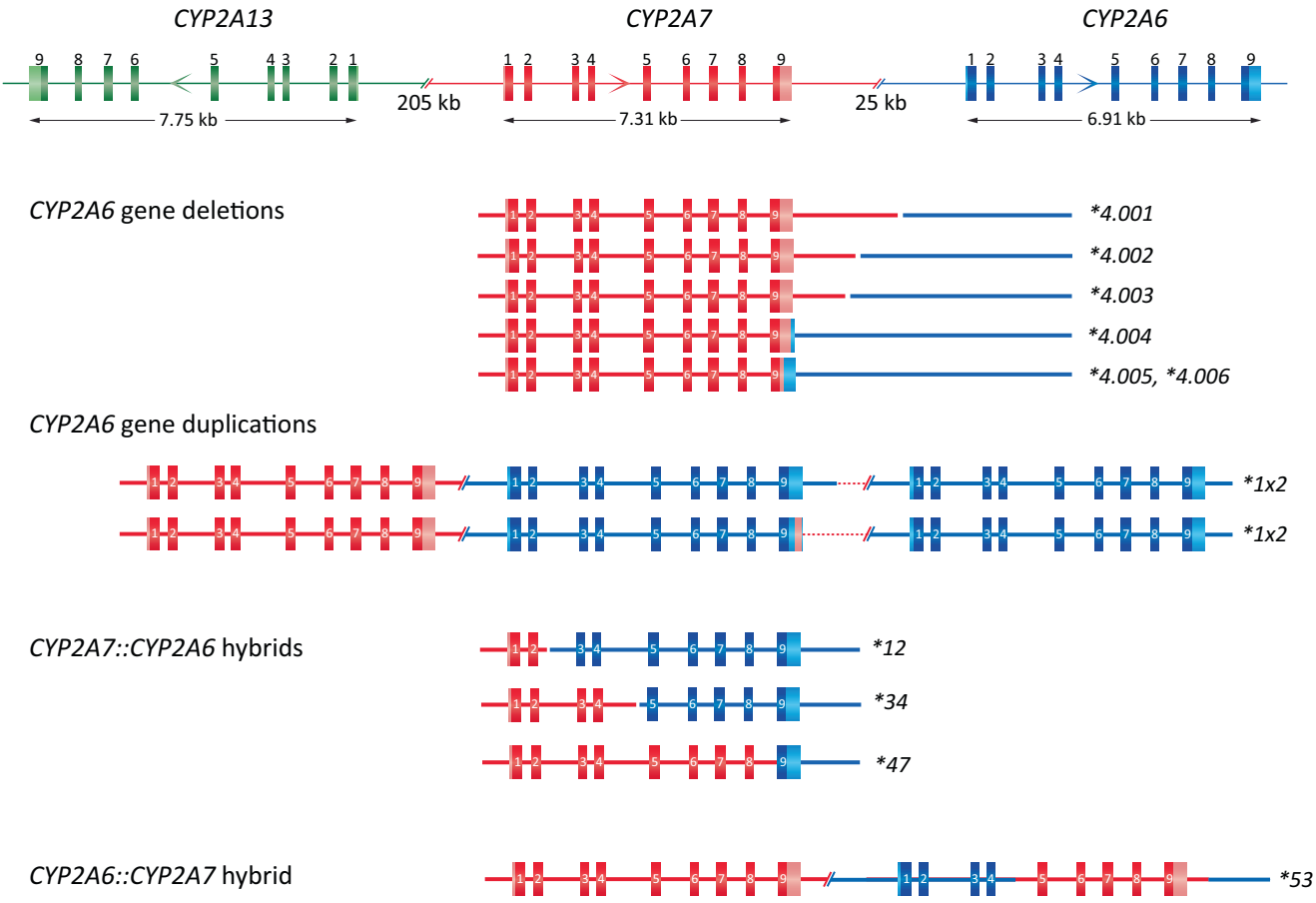

Figure 1 Overview of *CYP2A* gene locus and structural variation. The gene locus (top panel) spans approximately 252 kb and harbors three genes, *CYP2A6* (blue), *CYP2A7* (red), and *CYP2A13* (green). The locus is shown as the reverse complement; as indicated by the colored arrows, *CYP2A6* and *CYP2A7* are both encoded on the negative strand while *CYP2A13* is encoded on the positive strand. Numbered boxes depict exons. Deletion events with breakpoints downstream of *CYP2A7* or within its 3′-UTR created a series of alleles lacking the entire *CYP2A6* gene. Duplication events have been described for *CYP2A6*1*, which may contain *CYP2A7*-derived sequences as indicated by the red hashed lines. Three *CYP2A7::CYP2A6* hybrids (**12*, **34*, and **47*) have been identified to date with switches in intron 2, intron 4, and intron 8, respectively. The exact switch positions remain unknown due to high sequence homologies between the two genes. One *CYP2A6::CYP2A7* hybrid (**53*) has also been identified, which occurred in a duplication arrangement with an intact copy of *CYP2A7*. Additional and up-to-date information about SVs can be found in the Structural Variation document on the PharmVar *CYP2A6* gene page.[74]

limited number of star alleles have been defined for *CYP2A13*, including the common *CYP2A13*2* and **3* alleles. In candidate gene studies, the nonsynonymous SNVs present in *CYP2A13*2* and **3* have been associated with a decreased risk for head and neck, lung, and bladder cancers.[70] Star alleles for *CYP2A13* on PharmVar are reported against the updated RefSeq NG_007928.1; however, no additional curation has been undertaken, and no new alleles have been added since 2011. Thus, it is likely that both common and many rare alleles exist which are currently not catalogued by PharmVar.

THE PHARMVAR *CYP2A6* GENE PAGE

The PharmVar *CYP2A6* gene page presents all defined star alleles, along with other important information for each allele, such as (i) cross-reference with the legacy name(s) under which a particular allele was first published, (ii) SNVs defining the allele, (iii) allele evidence level, and (iv) references supporting allele definitions. The Compare View feature allows the user to toggle between the standard allele table and the graphical *C*omparative *A*llele *V*iew*E*r (CAVE) display. The *CYP2A6* gene page also includes Read Me, Change Log, and Structural Variation documents providing additional relevant gene-specific information and resources. Also, each haplotype has a unique numeric identifier, the PharmVar ID (PVID). PVIDs and their respective star allele (haplotype) definitions can be tracked in the database via the PVID Lookup function.

Variant mapping

SNVs are mapped to the genomic (NG_008377.1) and transcript (NM_000762.6) reference sequences, as well as the GRCh37 (NC_000019.9) and GRCh38 (NC_000019.10) genome builds. NM_000762.6 represents the Matched Annotation from NCBI and EMBL-EBI (MANE) select transcript which matches GRCh38 and is 100% identical with NG_008377.1 across the 5′UTR, coding regions, and 3′UTR.

PharmVar defines star alleles (haplotypes) for *CYP2A6* in the following regions: the 5′ flanking sequence, the 5′UTR, all nine exons, all eight introns, the 3′UTR, and 255 bp of the 3′ flanking

sequence. The 5′ and 3′ limits (start and end positions) correspond to NG_008377.1 positions 3721 through 11,907 counting from the sequence start and −1301 through 6886 counting from the ATG start codon. In this review (unless otherwise stated), we use positions on the genomic reference sequence NG_008377.1 counting from the ATG start codon.

The PharmVar *CYP2A6* gene page supports easy cross-referencing of SNVs across genomic and transcript reference sequences and the determination of the position of an SNV using two count modes (from the first nucleotide of the reference sequence or from the "A" of the ATG translation start codon). In addition to coordinates, the PharmVar display also provides SNV rsID numbers (if available) and corresponding amino acid changes. **Figure 2a** provides an excerpt of the *CYP2A6* gene page illustrating how information is displayed for allelic variants. Furthermore, in the Variant Window, SNV annotations are noted according to the Human Genome Variation Society (HGVS),[71] and in PharmVar-style format for all reference sequences side-by-side.

CYP2A6 haplotype evidence levels

PharmVar assigns "Allele Evidence Levels" for each star allele. As described in the Allele Designation and Evidence Criteria Level document,[72] definitive (Def), moderate (Mod), and limited (Lim) indicate the level of evidence supporting an allele definition (not their function). Numerous *CYP2A6* alleles are currently labeled as "Lim," as their definitions do not fulfill current PharmVar requirements (two of those are shown in **Figure 2a**), or because the phase of the SNVs was computationally inferred as opposed to experimentally determined. PharmVar solicits submissions for all star alleles with "Lim" and "Mod" evidence levels to substantiate their current definitions. In the effort to further corroborate haplotype-based star allele definitions, PharmVar also encourages encore submissions for star alleles with definitive evidence levels and single citations.

Core allele definitions

A core allele is defined only by sequence variants that cause an amino acid change or impact function by, for example, changing expression levels or interfering with splicing, and that are present in all haplotypes listed within an allele group with the same star number. These variants are also referred to as "core" variants or SNVs. With this rule-based system, alleles of a group (also referred to as suballeles) can be collapsed into a single "core" definition representing all haplotypes categorized under a star (*) number.[72,73] For example, all *CYP2A6*15* suballeles share -48T>G (impacting expression) and 2134A>G (p.K194E, rs199916117); thus, these two SNVs are part of the *CYP2A6*15* core allele definition (**Figure 2a,b**). For *CYP2A6*18*, only 5668A>T (p.Y392F, rs1809810) is shared among the three known suballeles, and therefore it is the only SNV of this particular core allele definition. Of note, *CYP2A6* core allele definitions are utilized for clinical annotations throughout PharmGKB.

All alleles under a star number are presumed to be functionally equivalent. Thus, even if a test is capable of distinguishing suballeles, these can be simply reported using core allele definitions.

The PharmVar Comparative Allele ViewEr and filter options

The Comparative Allele ViewEr (CAVE) tool allows users to compare core alleles graphically. **Figure 2c** visualizes the utility of this tool by comparing the *CYP2A6*9, *13, *15, *24, *26, *46* and **50* alleles. In this display mode, it is easy to see which SNVs alter function, which core SNVs are unique to a star allele (e.g., 2134A>G for *CYP2A6*15*), and which ones are shared among two or more star alleles (e.g., -48T>G). Alternatively, haplotypes can also be selectively displayed with the "Haplotype Filter" or searched by using SNV positions or their rsID.

REVISIONS AND CHANGES TO PREVIOUSLY PUBLISHED STAR ALLELE DEFINITIONS

Many star allele definitions were initially based on incomplete allele characterization, such as missing information for the 5′ flanking region, genotype analysis rather than sequencing and/or omission of reporting synonymous SNVs. After carefully reviewing the available literature, some allele definitions were corrected, revised, or removed. For example, 1165G>A (rs1226465361) was added to the *CYP2A6*25* allele definition after reviewing the original paper. This SNV appears to have been inadvertently omitted when *CYP2A6*25* was first defined. On the other hand, *CYP2A6*3* was retired as the methods may have led to nonspecific co-amplification and sequencing of *CYP2A7* and were thus deemed inadequate for cataloguing. The expert panel's in-depth literature review also led to the merging of several suballeles for *CYP2A6*46* (Table 1). For example, *CYP2A6*46.004* was merged with **46.002*, and *CYP2A6*46.014* was merged with **46.012*. Definitions of several star alleles remain incomplete; however, as indicated by an allele evidence level of "Lim." The majority of these have not been sequenced across the entire region as required by PharmVar and stated above. For a complete list of changes, see the Change Log document provided on the PharmVar *CYP2A6* gene page.[74]

GENOTYPE TO PHENOTYPE TRANSLATION

The NMR is advantageous in that it enables quantitative estimation of the effect of genetic variants/diplotypes on CYP2A6 activity *in vivo*. As the NMR is a continuous measure, studies have used different cut points to distinguish NMR groupings. There are also no consistent groupings of *CYP2A6* genotypes to translate into phenotype or metabolizer status, for example, slow vs. fast metabolizers or intermediate vs. normal metabolizers, though there is some overlap in which alleles are included in the different phenotype groups in the literature.

Individual studies have classified CYP2A6 activity dichotomously (e.g., slow vs. normal)[23,24,55] or into tertiles or quartiles based on ranked NMR activity[75,76]; these methodologies often result in differing CYP2A6 activity cut points to group individuals. For instance, in addition to NMR quartiles,[24–26,33] clinically relevant cut points associated with smoking cessation outcomes have also been used recently to dichotomize patients.[55,57]

CYP2A6 genotype has also been used to categorize participants. For example, several studies assigned *CYP2A6*1/*1* genotypes when individuals tested negative for any of the interrogated SNVs and classified these as having "normal" function. Individuals with

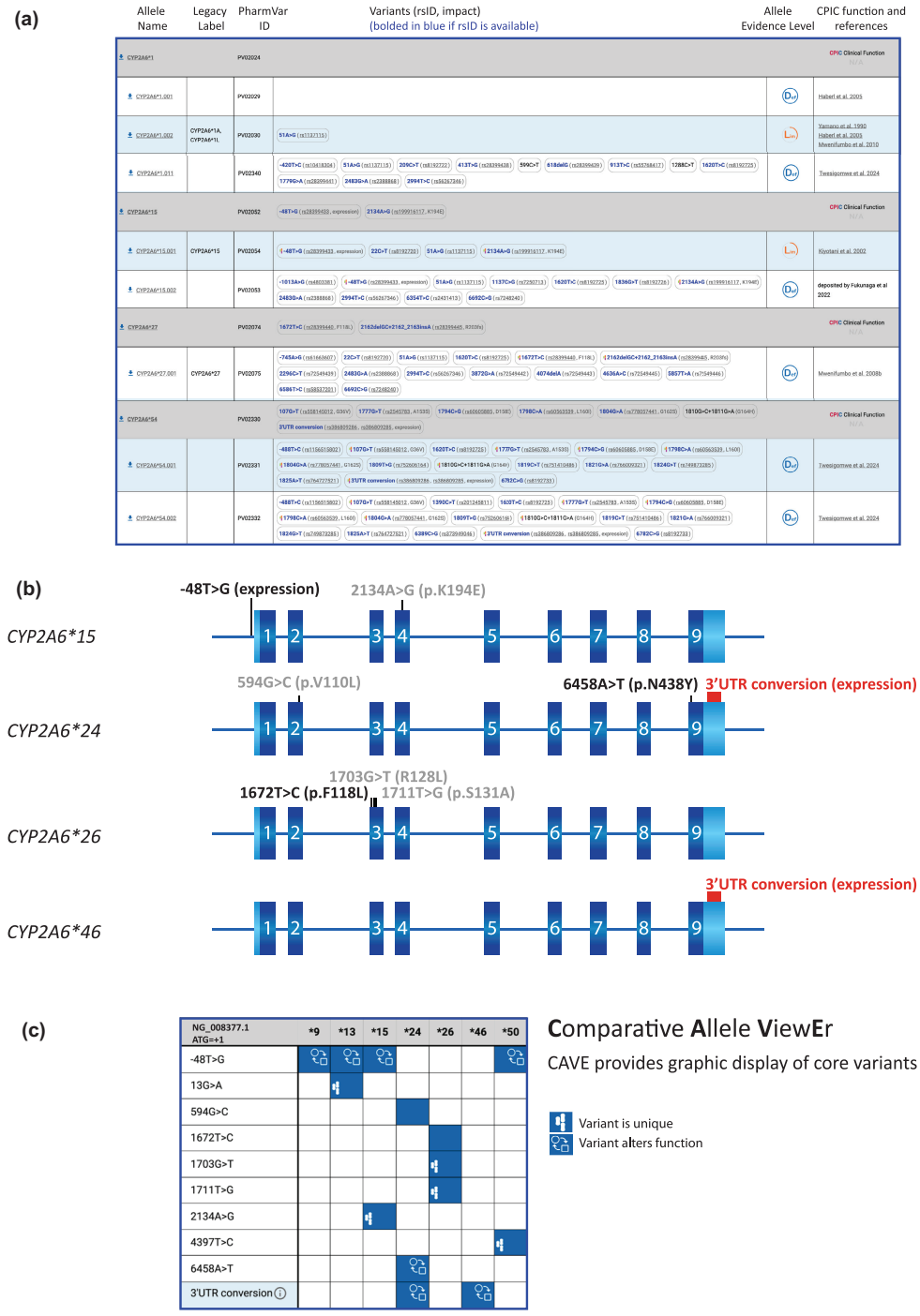

Figure 2 Overview of core alleles, suballeles, and the graphical Comparative Allele ViewEr (CAVE). An excerpt of the *CYP2A6* gene page is provided in panel (**a**). Gray bars show the *CYP2A62*1*, **15*, **27*, and **54* core allele definitions, whereas blue and white backgrounds signify suballeles. For *CYP2A6*1*, only selected suballeles are depicted, while all current suballeles are shown for **15*, **27*, and **54*. Positions correspond to NG_008377.1 as the reference sequence (counting from the translation start). Star allele names, PharmVar ID (PVID), allele evidence levels, and references are shown for each allele. Legacy allele names are cross-referenced. rsIDs (where existing) are provided for each variant with direct link-out capability to dbSNP. Of note, CPIC has not assigned clinical allele function for CYP2A6 and therefore, the function is shown as N/A for all star alleles. Panel (**b**) is a graphical representation of selected alleles with their respective core SNVs. Core SNVs are shared among all alleles listed under that star number and cause an amino acid change (*CYP2A6*15*, **24*, and **26*) or impact gene expression (**24* and **46*); those known to impact function are shown in bolded black font and those with unknown or uncertain function are in gray font. The *CYP2A7*-derived 3′UTR conversion (which also impacts gene expression) is highlighted in red. Blue numbered boxes represent exons and untranslated regions are shown in light blue. Panel (**c**) represents the graphical output of the Comparative Allele ViewEr (CAVE) that allows easy visualization of which SNVs are (i) unique to a star allele, for example, 13G>A (rs28399434, G5R) is only found on **13* and 2134A>G is only found on **15*, etc., or (ii) occur on two or more star alleles, for example, -48T>G is present on **9*, **13*, **15*, and **50*. The PharmVar symbols indicate which variants are unique to an allele and the function symbol signifies that the variant alters function.

Table 1 Overview of selected changes and revisions made to *CYP2A6* alleles as the gene was transferred into the PharmVar database or thereafter

Legacy allele designation	Change	Notes
3′ UTR	Revised	Revised the definition of the *CYP2A7*-derived 3′UTR conversion
3′ flanking	Removed	SNVs in the 3′ flanking region (past the 3′UTR conversion) are no longer part of *CYP2A6* star allele definitions
51G>A (rs1137115)	Updated	51G>A was revised to 51A>G due to the 'A' being the reference on NG_008377.1. Consequently, all alleles that had 'A' are now shown as reference and those with 'G' are shown as having an SNV. This change affected numerous alleles as the variant 'A' represents the more commonly found nucleotide
**1B1-*1B17*	Reassigned	Reassigned as **46* due to the presence of the 3′-UTR conversion which increases function; some suballeles were merged and several required corrections
**3*	Removed	Used inadequate methods; allele likely does not exist
**4C*	Merged	Literature review suggests that this allele matched **4A*
**12* suballeles	Removed	Due to no (**12A*) or limited (**12B* and **12C*) sequence information, differences among **12* allelic variants remain unclear; therefore, no suballele identifiers are provided at this time
**25*	Corrected	Added 1165G>A; this SNV was reported to be present but was inadvertently omitted in the legacy allele definition[98]
**39, *40, *43, *44 & *45*	Revised	Added intronic SNVs to respective allele definitions which are described by Piliguian.[99] These SNVs were omitted when alleles were first designated and described by Piliguian *et al*.[100]
**48-*52*	New star alleles	New star alleles found in Japanese populations (deposited by Fukunaga *et al*.)
**53*		First *CYP2A6::CYP2A7* hybrid described to date[64]
**54-*56*		New star alleles discovered in sub-Saharan African populations[83]

one *CYP2A6*1* allele and one copy of a decreased function allele (*CYP2A6*9* or *CYP2A6*12*) were classified as intermediate, and those with two copies of a decreased function or one or more copies of a nonfunctional allele (*CYP2A6*2* or *CYP2A6*4*) as being slow.[75,76] The relatively rare *CYP2A6*1x2* gene duplication has not been assessed extensively for increases in activity,[64,77,78] while *CYP2A6*46* has been associated with increased CYP2A6 activity in individuals of European-ancestry (consistent with *in vitro* work), but not in individuals of African-ancestry[67,79,80]; thus, at this time it is unclear whether either variant is consistently, and causally, related to increased CYP2A6 activity. Therefore, "rapid" or "ultrarapid" (or similar terms like "increased function" or "gain-of-function") classifications have not been used extensively. Finally, several studies have assessed nicotine metabolism using both the NMR and *CYP2A6* genotypes (for common star alleles) to guide/validate genotype-to-phenotype translations, categorizing participants trichotomously[27,81] or dichotomously.[82]

Recent genome-wide association studies of the NMR have identified non-star allele SNVs associated with increased function (e.g., rs56113850), which have been included in recent weighted genetic risk scores (wGRS).[55–58] These SNVs are intronic or proximal to the *CYP2A6* gene region, and some serve as proxies for functional star alleles; some may themselves have causal effects on gene regulation. Thus, unlike many CYP activity scoring approaches which only include star alleles (primarily amino acid changing SNVs), approximately 30% of CYP2A6 activity can be captured with the risk scores including GWAS SNVs (vs. 20% captured using common star alleles); however, these scores involve additional complexity as they include non-star allele SNVs (for which established genotyping procedures may not exist) and are currently ancestry-specific.

ALLELE FUNCTION

PharmVar has assembled a list of alleles with *in vivo* evidence of impact on nicotine metabolism (generally measured using the NMR) or coumarin metabolism (see table 3 in the Read Me document available at PharmVar[74]). To be listed, the functional impact of an allele must be supported by at least two publications.

CYP2A6 ALLELE, GENOTYPE, AND PHENOTYPE FREQUENCIES ACROSS POPULATIONS

There are substantial differences in the frequencies of *CYP2A6* star alleles across populations. One prominent example is the full gene deletion allele, *CYP2A6*4*. Note that *CYP2A6*47*, a *CYP2A7::CYP2A6* hybrid allele, has not been traditionally distinguished from *CYP2A6*4* due to its structural similarity and lack of activity, causing sources on *CYP2A6*4* frequency to include *CYP2A6*47* (see the *CYP2A6* Structural Variation document for more information). In European Ancestry populations (EUR), *CYP2A6*4* is rare at ~0.3%. It is also rare among Southwest and Northern Plains American Indian populations (0.3–1.6%), but is more common in African-American (AAC; 2.3%), Sub-Saharan African (SSA; 3.1%), and South Asian ancestry populations (SAS; 3.8%), and is rather frequent in East Asian

(EAS) and Alaska Native Yupik populations (e.g., > 20% in certain subpopulations).[6,64,83] *CYP2A6*12*, a *CYP2A7::CYP2A6* hybrid allele predicted to have severely decreased or no activity, is the most common SV in EUR (2.3% allele frequency), but is rare in other populations.[64,83] *CYP2A6*17*, a decreased function allele, is common in SSA and AAC populations but is absent or rare in EUR, SAS, or EAS.[6,60,83]

Recently, wGRS were constructed to predict CYP2A6 activity from genotype in EUR and AAC populations.[55,57] While *CYP2A6*4*, *CYP2A6*9*, and *CYP2A6*12* contributed to variation in CYP2A6 activity in both populations and thus were retained in both wGRSs, others were not. For example, the core *CYP2A6*2* SNV was retained only in the EUR wGRS, while the core SNVs for *CYP2A6*17*, **20*, **25*, **26*, **27*, and **35* were included only in the AAC wGRS. These differences are likely accounted for by relatively higher allele frequencies in one ancestral group vs. the other, as effects on CYP2A6 activity for all variants present in both populations were similar in magnitude and direction.[57]

Early studies of nicotine clearance through the administration of deuterium-labeled nicotine found that EUR and Latino (LAT) individuals have similar rates of nicotine metabolism, while both are significantly faster metabolizers than AAC and Chinese-American smokers[84,85]; later studies of the urinary NMR confirmed these trends, finding that EAS had the lowest NMR among ancestry groups studied.[86]

Another study of the urinary NMR (measured using $3'\text{-HC}_{\text{total}}/\text{cotinine}_{\text{free}}$ in regular smokers) within the multiethnic cohort study found significant differences in NMR across populations: EUR and AAC were not significantly different, Japanese and Native Hawaiians had significantly slower activity than EUR, and LAT had significantly faster activity than EUR.[87] Other studies have found overall lower mean NMRs in AAC compared to EUR, but no difference once individuals with known variants were excluded, implying that population differences are caused primarily by differences in *CYP2A6* star allele frequencies.[19,88] Lower activity among Japanese individuals is consistent with the high frequency of *CYP2A6*4* gene deletions in that population. On the other hand, Yupik Alaska Natives have substantially higher NMRs than AAC and EUR, even when comparisons are limited to *CYP2A6*1/*1* individuals, suggesting the presence of unknown variants (including those influencing regulation), environmental influences, or other causes.[88]

CYP2A6 ALLELE CHARACTERIZATION AND ALLELE CALLING

CYP2A6 genotyping

Due to the existence of the highly homologous *CYP2A7* and *CYP2A13* genes, *CYP2A6* genotype calls using more recently developed methods (including array genotyping, amplicon-based exon sequencing, and targeted capture NGS), are not always consistent with those used previously (e.g., nested PCR),[89] and are not always consistent between each other.[79] Specific positions within the gene are very likely to be genotyped discordantly between methods, especially in the exon 9 and/or 3′UTR regions, where, in European populations, approximately half of *CYP2A6* alleles possess a 3′UTR identical to *CYP2A7*,[79] that is, the *CYP2A7*-derived 3′UTR conversion. Thus, short read sequencing without a gene-specific protocol, that is, amplifying *CYP2A6* specifically and preventing read alignment to *CYP2A7* through reference genome masking, may not reliably produce high-quality data in exon 9 or the 3′UTR (**Figure 3**). There are other regions with high homology which may lead to genotyping errors for star alleles with core SNVs in exons 5 or 9, including *CYP2A6*5*, **7*, **8*, **11*, **24* and *CYP2A6*35*. Additionally, these regions may be challenging when identifying and characterizing novel haplotypes.

CYP2A6 SVs can be detected with quantitative PCR approaches employing commercially available TaqMan copy number assays, where *CYP2A6*1x2*, **4*, **12*, **34*, and **53* may be distinguished by targeting three different gene regions such as *CYP2A6* introns 1, 2, and 7.[64] Thus far, protocols have not been developed for routine genotyping of SVs using long-read sequencing. However, further progress in the use of long reads for *CYP2A6* sequencing should allow for the development of tools to facilitate accurate variant and haplotype calling, including SVs (similar to tools developed for *CYP2D6* by, e.g., Pacific Biosystems).[90]

Characterizing novel *CYP2A6* alleles

In the past, relatively low-throughput methods such as cloning of *CYP2A6*-specific XL-PCR products[14] or amplicon sequencing using a nested PCR protocol[57,91] were used to characterize most of the alleles currently catalogued by PharmVar, including the more recently deposited haplotypes. Given the aforementioned challenges posed by short read methods, the use of single molecule, long read sequencing represents an alternative method allowing for high(er) throughput and accurate characterization of *CYP2A6* alleles. Indeed, several novel alleles were deposited in PharmVar after fully phased sequences of the entire *CYP2A6* gene (including introns and up/downstream regions) were obtained using Pacific Biosciences (Menlo Park, CA) single molecule sequencing.[83] Specifically, long range PCR products encompassing the entire *CYP2A6* gene, upstream sequences and the 3′UTR region were amplified and subjected to Pacific Biosciences High Fidelity (HiFi) sequencing. This approach enables the detection and accurate phasing of all SNVs on both alleles. For one of the participants, this analysis revealed *CYP2A6*17.002* and a novel haplotype, *CYP2A6*54.001*. Of note, the latter contains the 3′UTR conversion which was fully resolved as shown in **Figure 4c, d**.[83] Furthermore, *CYP2A6*54* was found to have a 49-bp long region in exon 3/intron 3 corresponding to *CYP2A7* (this feature is listed in the Structural Variation document as "ex3/in3 conversion"). The majority of the SNVs in this region were missed initially when analyzing WGS short read data and were only detected and phased with HiFi long read technology. Other targeted long-read pharmacogene sequencing platforms/resources, such as the Twist Alliance Long-Read Pharmacogenetics Panel (Twist Bioscience; South San Francisco, CA) or Oxford Nanopore (Oxford Nanopore Technologies; Oxford, UK) single molecule sequencing may also be applied for *CYP2A6* genotyping and allele characterization.[92] However, there are currently no published reports that describe the use of these approaches for *CYP2A6*. The main limitation to the universal use of long read sequencing

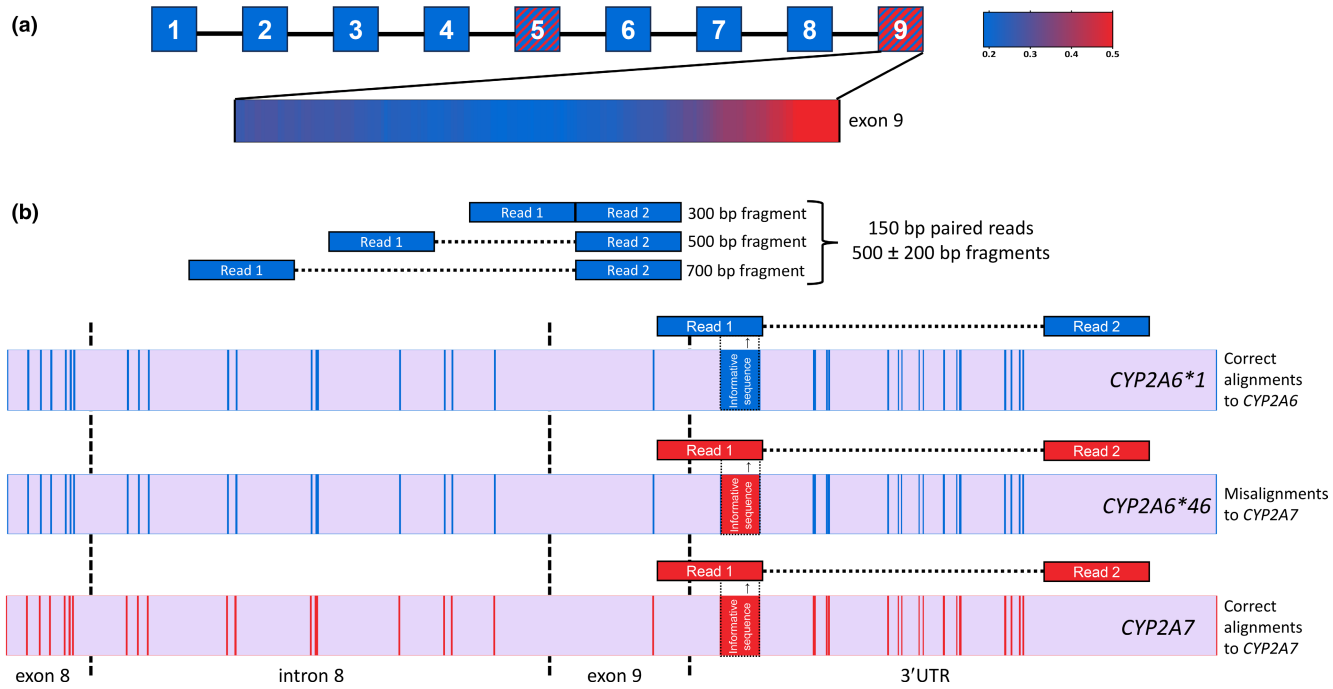

Figure 3 NGS with paired-end short reads can produce misalignments due to high sequence identity between *CYP2A6* and *CYP2A7*. Panel (**a**) visualizes the results of a simulation analysis in which 2×150 bp paired reads with average fragment sizes of 500 bp were generated from the *CYP2A6* RefSeq, with an average coverage per position of 100 reads (*CYP2A6* exons/introns are not drawn to scale); a second simulation used reads into which the 3′UTR conversion was introduced to simulate *CYP2A6*46*. For each position, the percentage of total reads that could not be disambiguated between *CYP2A6* or *CYP2A7*, or which misaligned to *CYP2A7* (in the case of the simulated *CYP2A6*46*), was determined. Misalignments were common in exon 5 and, when the 3′UTR conversion was present, exon 9. Misalignments in exon 9 are shown by position in a heatmap where the darkest red corresponds to >50% of misaligned reads. Panel (**b**) depicts a region from exon 8 to the 3′UTR (to scale) demonstrating how read misalignments can occur when NGS short reads have a mean fragment size of 500 bp. Read pairs are displayed as boxes with a connecting hashed black line. Blue and red lines indicate positions where *CYP2A6* and *CYP2A7* differ, that is, represent informative positions, and the large blue and red block indicates the 3′UTR conversion region. Fragment size variation can produce misalignments when a read overlaps the 3′UTR conversion, as demonstrated by a hypothetical read pair shown above the *CYP2A6*46* graph, which would produce "unambiguous" alignments to *CYP2A7*.

for *CYP2A6* allele characterization is the relatively higher cost compared to other technologies, especially in resource-limited settings. A CRISPR-Cas9-based approach that was developed to characterize the entire *CYP2D* locus could also be harnessed to characterize particularly complex structural variants/rearrangements of the *CYP2A* locus.[93]

Tools to facilitate *CYP2A6* star allele calling

The strengths and limitations of the methodologies described above imply that accurate *CYP2A6* allele characterization often requires a combination of approaches. Given the plethora of emerging small and large datasets globally, bioinformatics algorithms for calling *CYP2A6* star alleles from NGS datasets are needed for data interpretation. At present, existing tools that support *CYP2A6* allele detection from NGS data include StellarPGx, Aldy, and PyPGx.[94–96] At present, all three tools support high-coverage short read WGS data as their main input, while their respective versions supporting long read and other data are under development or being further enhanced. The analysis of data other than WGS (e.g., genome-wide arrays, WES, custom NGS gene panels, or sequencing of XL-PCR amplicons) using these tools may need to be interpreted with caution as input data may not yield accurate haplotypes and/or SV calls, especially for cases with complex diplotypes. In comparison to short read-based NGS, long reads are considerably less prone to read misalignments, which is highly advantageous for characterizing *CYP2A6* star alleles using these computational algorithms. However, calling *CYP2A6* star alleles from long read sequencing data presents different computational challenges compared to those from short read data. For example, when inferring the presence of SVs, read depth-based approaches tend to be less effective for long read data given their typically lower read depth compared to short read data. Consequently, the detection of SVs from long read data may require various nuanced solutions. Notably, typical long read fragments can span the entire length of *CYP2A6*, which facilitates the implementation of read-based phasing strategies such as WhatsHap[97] within existing allele callers, rather than the statistical phasing-based or combinatorial-based diplotype assignment strategies used for short read data.[94–96] Overall, the use of tools such as StellarPGx, Aldy, and PyPGx provides the opportunity to perform *CYP2A6* pharmacogenomic analysis within large-scale short read datasets in research settings. Novel haplotypes identified with these tools from short read NGS will require validation to be considered by PharmVar for naming; the necessity of validating novel haplotypes is underscored in

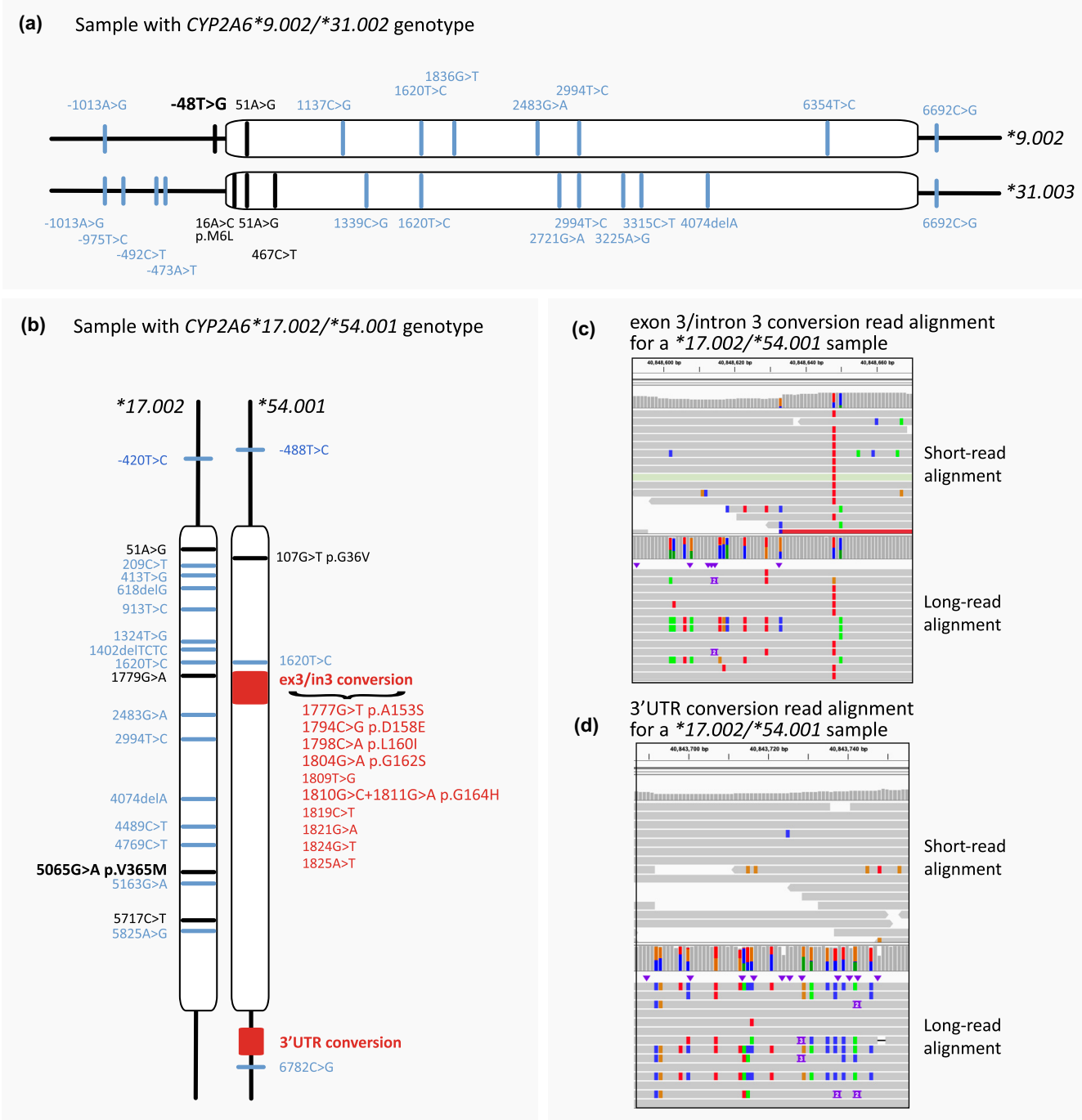

Figure 4 *CYP2A6* haplotype characterization using single molecule sequencing. Panels (**a**) and (**b**) showcase two samples for which both haplotypes were initially found using WGS and subsequently further characterized with single molecule long read sequencing. *CYP2A7*-derived sequences are highlighted in red. SNVs in coding regions are in black font, and those in noncoding regions are shown in blue font. SNVs known to alter function are highlighted in larger font. The *CYP2A6*9.002* allele in (**a**) represents a fully characterized sequence that was initially designated as **9B*; several SNVs have since been added to the allele's definition. The subject's second allele represents a new *CYP2A6*31* suballele, **31.003*. The *CYP2A6*17.002* and **54.001* alleles found in the subject shown in (**b**) are both novel. *CYP2A6*54.001* has a 49-bp long region spanning parts of exon 3 and intron 3 (ex3/in3) that is *CYP2A7*-derived and listed as a conversion in the Structural Variation document. This region contains 11 SNVs, of which six contribute to five amino acid changes. *CYP2A6*54.001* also has the *CYP2D7*-derived 3′UTR conversion. The top portions of panels (**c**) and (**d**) visualize the challenge of resolving *CYP2A7* conversion regions with WGS short read sequencing. As shown in their respective bottom portions, these challenging regions can be resolved using single molecule long read sequencing. Screenshots of read alignments represent the sample depicted in panel (**b**).

the next section on several examples which illustrate the challenges of calling *CYP2A6* alleles from WGS short read data.

Examples of *CYP2A6* characterization challenges

As of the writing of this review, PharmVar does not accept submissions that are solely based on short read WGS, that is, which have not been validated with another method. This is largely due to read alignment and SNV/SV detection issues arising from the high homology between *CYP2A6* and *CYP2A7*, in addition to the phasing limitations posed by short read data. The recent characterization of *CYP2A6*54* provides an example where SNVs in two *CYP2A7*-derived regions, in the 3′UTR and exon 3/intron 3 were not called from high-coverage short read WGS data; instead, read misalignments to *CYP2A7* caused insufficient allele depths to support the correct variant calls in *CYP2A6*. Since HiFi long read technology compensates for alignment penalties across the entire *CYP2A6* gene region, long reads allowed for the detection of all SNVs, in addition to phasing and unequivocal detection of both conversion events (**Figure 4c,d**). Given that the *CYP2A6* 3′UTR conversion is highly prevalent (*CYP2A6*46* alone is estimated to range from 6% to 42% globally),[83] it is imperative that all *CYP2A6* submissions to PharmVar have accurately ascertained the presence and phase of all SNVs, including those in conversions.

As mentioned, SVs are difficult to assess accurately with short read WGS alone, but long read sequencing protocols may also result in ambiguity. For example, the submission for *CYP2A6*56.001* provided evidence from short read WGS data indicating the presence of a duplication. Two gene copies in the sample, the novel *CYP2A6*56.001* allele and the **35.003* suballele, were fully characterized via targeted Pacific Biosciences HiFi resequencing. However, the duplicated gene copy eluded characterization as it was not captured when XL-PCR sequencing templates were generated as primers were not designed to amplify duplicated gene copies. Using only the short read data, it was inconclusive as to whether the sample had a *CYP2A6*35.003* or a **56.001* duplication. It also remains unclear whether the duplicated gene copies were 100% identical. Nevertheless, short read WGS may be sufficient for calling *CYP2A6* alleles if (i) the data meets quality metrics (e.g., read depth > 30x, variants pass filters such as Quality by Depth and Allele Balance), (ii) there are no SVs, and preferably (iii) there is additional support from inheritance/trios that substantiates the PharmVar catalogued haplotypes. As additional information regarding the accuracy of haplotypes that are characterized solely by short read WGS data, especially those in copy-neutral, conversion-negative samples becomes available, PharmVar will adapt/update submission requirements. For the time-being, however, novel haplotypes discovered by short read WGS must be validated by appropriate orthogonal methods/approaches.

CONCLUSIONS

This PharmVar *CYP2A6* GeneFocus paper sheds light on this highly polymorphic gene to facilitate further work investigating its clinical impact. We present an overview of current knowledge, examples of past methods for allele characterization, and several obstacles limiting the full understanding of this complex genetic region, which may be addressed by modern genetic approaches.

FUNDING

This work was funded by a Canadian Institutes of Health Research (CIHR) Project grant (PJY-159710; M.J.C. and R.F.T.) and Canadian Institutes of Health Research (CIHR) Foundation grant (FDN-154294; R.F.T.), and a Canada Research Chair in Pharmacogenomics (R.F.T.). We also acknowledge funding from the CIHR Canadian Graduate Scholarships – Doctoral Program (A.W.R.L.). PharmGKB and CPIC (T.E.K., M.W.C., and K.S.) are supported by the National Institutes of Health NHGRI/NICHD/NIDA U24 HG010615 and NHGRI U24 HG010135. D.T. is supported by funding from the Sydney Brenner Charitable Trust, and his postdoctoral fellowship is hosted at the University of the Witwatersrand in partnership with the University of Edinburgh. PharmVar is fully supported by the Children's Mercy Research Institute (CMRI).

CONFLICT OF INTEREST

C.N. and G.S. are employed by PharmGenetix GmbH, a private laboratory providing PGx testing, reporting, and interpretation services. All other authors declared no competing interests for this work.

1. Hukkanen, J., Jacob, P. & Benowitz, N.L. Metabolism and disposition kinetics of nicotine. *Pharmacol. Rev.* **57**, 79–115 (2005).
2. Nagata, K., Matsunaga, T., Gillette, J., Gelboin, H.V. & Gonzalez, F.J. Rat testosterone 7 alpha-hydroxylase. Isolation, sequence, and expression of cDNA and its developmental regulation and induction by 3-methylcholanthrene. *J Biol Chem.* **262**, 2787–2793 (1987).
3. Yamano, S., Tatsuno, J. & Gonzalez, F.J. The CYP2A3 gene product catalyzes coumarin 7-hydroxylation in human liver microsomes. *Biochemistry.* **29**, 1322–1329 (1990).
4. Fernandez-Salguero, P. *et al*. A genetic polymorphism in coumarin 7-hydroxylation: sequence of the human CYP2A genes and identification of variant CYP2A6 alleles. *Am. J. Hum. Genet.* **57**, 651–660 (1995).
5. von Weymarn, L.B., Retzlaff, C. & Murphy, S.E. CYP2A6- and CYP2A13-catalyzed metabolism of the nicotine Δ5'(1′)iminium ion. *J. Pharmacol. Exp. Ther.* **343**, 307–315 (2012).
6. Tanner, J.A. & Tyndale, R.F. Variation in CYP2A6 activity and personalized medicine. *J Pers Med.* **7**, 7–16 (2017). https://doi.org/10.3390/jpm7040018.
7. Nakajima, M. *et al*. Role of human cytochrome P4502A6 in C-oxidation of nicotine. *Drug Metab. Dispos.* **24**, 1212–1217 (1996).
8. Messina, E.S., Tyndale, R.F. & Sellers, E.M. A major role for CYP2A6 in nicotine C-oxidation by human liver microsomes. *J. Pharmacol. Exp. Ther.* **282**, 1608–1614 (1997).
9. Oscarson, M. Genetic polymorphisms in the cytochrome P450 2A6 (CYP2A6) gene: implications for interindividual differences in nicotine metabolism. *Drug Metab. Dispos.* **29**, 91–95 (2001).
10. Pianezza, M.L., Sellers, E.M. & Tyndale, R.F. Nicotine metabolism defect reduces smoking. *Nature* **393**, 750 (1998).
11. Wong, H.L., Murphy, S.E. & Hecht, S.S. Cytochrome P450 2A-catalyzed metabolic activation of structurally similar carcinogenic nitrosamines: N'-nitrosonornicotine enantiomers, N-nitrosopiperidine, and N-nitrosopyrrolidine. *Chem. Res. Toxicol.* **18**, 61–69 (2005).

12. Patten, C.J., Smith, T.J., Friesen, M.J., Tynes, R.E., Yang, C.S. & Murphy, S.E. Evidence for cytochrome P450 2A6 and 3A4 as major catalysts for N'-nitrosonornicotine alpha-hydroxylation by human liver microsomes. *Carcinogenesis* **18**, 1623–1630 (1997).
13. Benowitz, N.L., Hukkanen, J. & Jacob, P. 3rd. Nicotine chemistry, metabolism, kinetics and biomarkers. *Handb. Exp. Pharmacol.* **192**, 29–60 (2009).
14. Mwenifumbo, J.C., Zhou, Q., Benowitz, N.L., Sellers, E.M. & Tyndale, R.F. New CYP2A6 gene deletion and conversion variants in a population of Black African descent. *Pharmacogenomics* **11**, 189–198 (2010).
15. Dempsey, D. *et al*. Nicotine metabolite ratio as an index of cytochrome P450 2A6 metabolic activity. *Clin. Pharmacol. Ther.* **76**, 64–72 (2004).
16. Giratallah, H.K. *et al*. Nicotine metabolite ratio: comparison of the three urinary versions to the plasma version and nicotine clearance in three clinical studies. *Drug Alcohol Depend.* **223**, 108708 (2021).
17. St Helen, G. *et al*. Reproducibility of the nicotine metabolite ratio in cigarette smokers. *Cancer Epidemiol. Biomarkers Prev.* **21**, 1105–1114 (2012).
18. Tanner, J.A., Chenoweth, M.J. & Tyndale, R.F. Pharmacogenetics of nicotine and associated smoking behaviors. *Curr. Top. Behav. Neurosci.* **23**, 37–86 (2015).
19. Chenoweth, M.J. *et al*. Known and novel sources of variability in the nicotine metabolite ratio in a large sample of treatment-seeking smokers. *Cancer Epidemiol. Biomarkers Prev.* **23**, 1773–1782 (2014).
20. Wassenaar, C.A. *et al*. CYP2A6 reduced activity gene variants confer reduction in lung cancer risk in African American smokers – findings from two independent populations. *Carcinogenesis* **36**, 99–103 (2015).
21. Wassenaar, C.A., Dong, Q., Wei, Q., Amos, C.I., Spitz, M.R. & Tyndale, R.F. Relationship between CYP2A6 and CHRNA5-CHRNA3-CHRNB4 variation and smoking behaviors and lung cancer risk. *J. Natl. Cancer Inst.* **103**, 1342–1346 (2011).
22. Giratallah, H. *et al*. CYP2A6 associates with respiratory disease risk and younger age of diagnosis: a phenome-wide association Mendelian randomization study. *Hum. Mol. Genet.* **33**, 198–210 (2024).
23. Lerman, C. *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.* **3**, 131–138 (2015).
24. Lerman, C. *et al*. Nicotine metabolite ratio predicts efficacy of transdermal nicotine for smoking cessation. *Clin. Pharmacol. Ther.* **79**, 600–608 (2006).
25. Lerman, C. *et al*. Genetic variation in nicotine metabolism predicts the efficacy of extended-duration transdermal nicotine therapy. *Clin. Pharmacol. Ther.* **87**, 553–557 (2010).
26. Patterson, F. *et al*. Toward personalized therapy for smoking cessation: a randomized placebo-controlled trial of bupropion. *Clin. Pharmacol. Ther.* **84**, 320–325 (2008).
27. Malaiyandi, V., Lerman, C., Benowitz, N.L., Jepson, C., Patterson, F. & Tyndale, R.F. Impact of CYP2A6 genotype on pretreatment smoking behaviour and nicotine levels from and usage of nicotine replacement therapy. *Mol. Psychiatry* **11**, 400–409 (2006).
28. Weinberger, A.H. *et al*. Cigarette use among individuals with alcohol use disorders in the United States, 2002 to 2016: trends overall and by race/ethnicity. *Alcohol. Clin. Exp. Res.* **43**, 79–90 (2019).
29. Gubner, N.R. *et al*. Cessation of alcohol consumption decreases rate of nicotine metabolism in male alcohol-dependent smokers. *Drug Alcohol Depend.* **163**, 157–164 (2016).
30. Ande, A., McArthur, C., Kumar, A. & Kumar, S. Tobacco smoking effect on HIV-1 pathogenesis: role of cytochrome P450 isozymes. *Expert Opin. Drug Metab. Toxicol.* **9**, 1453–1464 (2013).
31. Helleberg, M. *et al*. Mortality attributable to smoking among HIV-1-infected individuals: a nationwide, population-based cohort study. *Clin. Infect. Dis.* **56**, 727–734 (2013).
32. Bien-Gund, C.H. *et al*. Brief Report: Nicotine metabolism ratio increases in HIV-positive smokers on effective antiretroviral therapy: a cohort study. *J. Acquir. Immune Defic. Syndr.* **89**, 428–432 (2022).
33. Ashare, R.L. *et al*. Differences in the rate of nicotine metabolism among smokers with and without HIV. *AIDS* **33**, 1083–1088 (2019).
34. Arens, Y. *et al*. Brief Report: HIV infection does not explain higher nicotine metabolism in people living with HIV. *J. Acquir. Immune Defic. Syndr.* **91**, 497–501 (2022).
35. Medaglio, D. *et al*. Nicotine metabolite ratio decreases after switching off efavirenz-based therapy in people with HIV who smoke. *Clin. Pharmacol. Ther.* **115**, 80–85 (2024).
36. Metzger, I.F., Dave, N., Kreutz, Y., Lu, J.B.L., Galinsky, R.E. & Desta, Z. CYP2B6 genotype-dependent inhibition of CYP1A2 and induction of CYP2A6 by the antiretroviral drug efavirenz in healthy volunteers. *Clin. Transl. Sci.* **12**, 657–666 (2019).
37. Park, S.R. *et al*. CYP2A6 and ERCC1 polymorphisms correlate with efficacy of S-1 plus cisplatin in metastatic gastric cancer patients. *Br. J. Cancer* **104**, 1126–1134 (2011).
38. Jeong, J.H. *et al*. Associations between CYP2A6 polymorphisms and outcomes of adjuvant S-1 chemotherapy in patients with curatively resected gastric cancer. *Gastric Cancer* **20**, 146–155 (2017).
39. Al Koudsi, N., Hoffmann, E.B., Assadzadeh, A. & Tyndale, R.F. Hepatic CYP2A6 levels and nicotine metabolism: impact of genetic, physiological, environmental, and epigenetic factors. *Eur. J. Clin. Pharmacol.* **66**, 239–251 (2010).
40. Dempsey, D., Jacob, P. 3rd & Benowitz, N.L. Nicotine metabolism and elimination kinetics in newborns. *Clin. Pharmacol. Ther.* **67**, 458–465 (2000).
41. Molander, L., Hansson, A. & Lunell, E. Pharmacokinetics of nicotine in healthy elderly people. *Clin. Pharmacol. Ther.* **69**, 57–65 (2001).
42. Higashi, E. *et al*. Human CYP2A6 is induced by estrogen via estrogen receptor. *Drug Metab. Dispos.* **35**, 1935–1941 (2007).
43. Tanner, J.A. *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.* **360**, 129–139 (2017).
44. Mwenifumbo, J.C., Sellers, E.M. & Tyndale, R.F. Nicotine metabolism and CYP2A6 activity in a population of Black African descent: impact of gender and light smoking. *Drug Alcohol Depend.* **89**, 24–33 (2007).
45. Benowitz, N.L., Lessov-Schlaggar, C.N., Swan, G.E. & Jacob, P. 3rd. Female sex and oral contraceptive use accelerate nicotine metabolism. *Clin. Pharmacol. Ther.* **79**, 480–488 (2006).
46. Matt, G.E. *et al*. Evaluation of urinary trans-3′-hydroxycotinine as a biomarker of children's environmental tobacco smoke exposure. *Biomarkers* **11**, 507–523 (2006).
47. Taghavi, T., Arger, C.A., Heil, S.H., Higgins, S.T. & Tyndale, R.F. Longitudinal influence of pregnancy on nicotine metabolic pathways. *J. Pharmacol. Exp. Ther.* **364**, 238–245 (2018).
48. Arger, C.A., Taghavi, T., Heil, S.H., Skelly, J., Tyndale, R.F. & Higgins, S.T. Pregnancy-induced increases in the nicotine metabolite ratio: examining changes during antepartum and postpartum. *Nicotine Tob. Res.* **21**, 1706–1710 (2019).
49. Schock, H. *et al*. Hormone concentrations throughout uncomplicated pregnancies: a longitudinal study. *BMC Pregnancy Childbirth* **16**, 146 (2016).
50. Choi, S.Y., Koh, K.H. & Jeong, H. Isoform-specific regulation of cytochromes P450 expression by estradiol and progesterone. *Drug Metab. Dispos.* **41**, 263–269 (2013).
51. Claire, R. *et al*. Pharmacological interventions for promoting smoking cessation during pregnancy. *Cochrane Database Syst. Rev.* **3**, CD010078 (2020).
52. Sellers, E.M., Kaplan, H.L. & Tyndale, R.F. Inhibition of cytochrome P450 2A6 increases nicotine's oral bioavailability and decreases smoking. *Clin. Pharmacol. Ther.* **68**, 35–43 (2000).
53. Benowitz, N.L. & Jacob, P. 3rd. Effects of cigarette smoking and carbon monoxide on nicotine and cotinine metabolism. *Clin. Pharmacol. Ther.* **67**, 653–659 (2000).
54. Lee, B.L., Benowitz, N.L. & Jacob, P. 3rd. Influence of tobacco abstinence on the disposition kinetics and effects of nicotine. *Clin. Pharmacol. Ther.* **41**, 474–479 (1987).

55. El-Boraie, A. *et al*. Evaluation of a weighted genetic risk score for the prediction of biomarkers of CYP2A6 activity. *Addict. Biol.* **25**, e12741 (2020).
56. Patel, Y.M. *et al*. Novel association of genetic markers affecting CYP2A6 activity and lung cancer risk. *Cancer Res* **76**, 5768–5776 (2016).
57. El-Boraie, A. *et al*. Transferability of ancestry-specific and cross-ancestry CYP2A6 activity genetic risk scores in African and European populations. *Clin. Pharmacol. Ther.* **110**, 975–985 (2020).
58. Buchwald, J. *et al*. Genome-wide association meta-analysis of nicotine metabolism and cigarette consumption measures in smokers of European descent. *Mol. Psychiatry* **26**, 2212–2223 (2021).
59. Xu, C. *et al*. An in vivo pilot study characterizing the new CYP2A6*7, *8, and *10 alleles. *Biochem. Biophys. Res. Commun.* **290**, 318–324 (2002).
60. Ho, M.K., Mwenifumbo, J.C., Zhao, B., Gillam, E.M. & Tyndale, R.F. 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* **18**, 67–75 (2008).
61. Stancil, S.L. *et al*. Evaluating metronidazole as a novel, safe CYP2A6 phenotyping probe in healthy adults. *Br. J. Clin. Pharmacol.* **85**, 960–969 (2019).
62. Whirl-Carrillo, M. *et al*. An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* **110**, 563–572 (2021).
63. Hoffman, S.M., Nelson, D.R. & Keeney, D.S. Organization, structure and evolution of the CYP2 gene cluster on human chromosome 19. *Pharmacogenetics* **11**, 687–698 (2001).
64. Langlois, A.W.R. *et al*. Genotyping, characterization, and imputation of known and novel CYP2A6 structural variants using SNP array data. *J. Hum. Genet.* **68**, 533–541 (2023).
65. Pitarque, M., von Richter, O., Oke, B., Berkkan, H., Oscarson, M. & Ingelman-Sundberg, M. Identification of a single nucleotide polymorphism in the TATA box of the CYP2A6 gene: impairment of its promoter activity. *Biochem. Biophys. Res. Commun.* **284**, 455–460 (2001).
66. Yoshida, R., Nakajima, M., Nishimura, K., Tokudome, S., Kwon, J.T. & Yokoi, T. Effects of polymorphism in promoter region of human CYP2A6 gene (CYP2A6*9) on expression level of messenger ribonucleic acid and enzymatic activity in vivo and in vitro. *Clin. Pharmacol. Ther.* **74**, 69–76 (2003).
67. Wang, J., Pitarque, M. & Ingelman-Sundberg, M. 3'-UTR polymorphism in the human CYP2A6 gene affects mRNA stability and enzyme expression. *Biochem. Biophys. Res. Commun.* **340**, 491–497 (2006).
68. Christian, K., Lang, M., Maurel, P. & Raffalli-Mathieu, F. Interaction of heterogeneous nuclear ribonucleoprotein A1 with cytochrome P450 2A6 mRNA: implications for post-transcriptional regulation of the CYP2A6 gene. *Mol. Pharmacol.* **65**, 1405–1414 (2004).
69. Su, T., Bao, Z., Zhang, Q.Y., Smith, T.J., Hong, J.Y. & Ding, X. Human cytochrome P450 CYP2A13: predominant expression in the respiratory tract and its high efficiency metabolic activation of a tobacco-specific carcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Cancer Res.* **60**, 5074–5079 (2000).
70. Vrzal, R. Genetic and enzymatic characteristics of CYP2A13 in relation to lung damage. *Int. J. Mol. Sci.* **22**, 12306 (2021).
71. Sequence Variant Nomenclature. Human Genome Variation Society <https://varnomen.hgvs.org/> Accessed January 11, 2024.
72. PharmVar. Allele Designation and Evidence Level Criteria document <www.pharmvar.org/documents/Allele_Designation_and_Evidence_Level_Criteria_V2.1.pdf>.
73. Gaedigk, A. *et al*. The evolution of PharmVar. *Clin. Pharmacol. Ther.* **105**, 29–32 (2019).
74. CYP2A6. Pharmacogene Variation Consortium <https://www.pharmvar.org/gene/CYP2A6>.
75. Chenoweth, M.J., O'Loughlin, J., Sylvestre, M.P. & Tyndale, R.F. CYP2A6 slow nicotine metabolism is associated with increased quitting by adolescent smokers. *Pharmacogenet. Genomics* **23**, 232–235 (2013).
76. Strasser, A.A., Malaiyandi, V., Hoffmann, E., Tyndale, R.F. & Lerman, C. An association of CYP2A6 genotype and smoking topography. *Nicotine Tob. Res.* **9**, 511–518 (2007).
77. Rao, Y. *et al*. Duplications and defects in the CYP2A6 gene: identification, genotyping, and in vivo effects on smoking. *Mol. Pharmacol.* **58**, 747–755 (2000).
78. Fukami, T., Nakajima, M., Yamanaka, H., Fukushima, Y., McLeod, H.L. & Yokoi, T. A novel duplication type of CYP2A6 gene in African-American population. *Drug Metab. Dispos.* **35**, 515–520 (2007).
79. Langlois, A.W.R. *et al*. Accuracy and applications of sequencing and genotyping approaches for CYP2A6 and homologous genes. *Pharmacogenet Genomics.* **32**, 159–172 (2022).
80. Bloom, J. *et al*. The contribution of common CYP2A6 alleles to variation in nicotine metabolism among European-Americans. *Pharmacogenet. Genomics* **21**, 403–416 (2011).
81. Verde, Z. *et al*. Effects of cigarette smoking and nicotine metabolite ratio on leukocyte telomere length. *Environ. Res.* **140**, 488–494 (2015).
82. Liakoni, E. *et al*. Effects of nicotine metabolic rate on withdrawal symptoms and response to cigarette smoking after abstinence. *Clin. Pharmacol. Ther.* **105**, 641–651 (2019).
83. Twesigomwe, D. *et al*. Characterization of CYP2B6 and CYP2A6 pharmacogenetic variation in sub-Saharan African populations. *Clin. Pharmacol. Ther.* **115**, 576–594 (2024).
84. Pérez-Stable, E.J., Herrera, B., Jacob, P. & Benowitz, N.L. Nicotine metabolism and intake in black and white smokers. *JAMA* **280**, 152–156 (1998).
85. Benowitz, N.L., Pérez-Stable, E.J., Herrera, B. & Jacob, P. Slower metabolism and reduced intake of nicotine from cigarette smoking in Chinese-Americans. *J. Natl Cancer Inst.* **94**, 108–115 (2002).
86. Kandel, D.B., Hu, M.C., Schaffran, C., Udry, J.R. & Benowitz, N.L. Urine nicotine metabolites and smoking behavior in a multiracial/multiethnic national sample of young adults. *Am. J. Epidemiol.* **165**, 901–910 (2007).
87. Park, S.L. *et al*. Genetic determinants of CYP2A6 activity across racial/ethnic groups with different risks of lung cancer and effect on their smoking intensity. *Carcinogenesis* **37**, 269–279 (2016).
88. Binnington, M.J. *et al*. CYP2A6 and CYP2B6 genetic variation and its association with nicotine metabolism in South Western Alaska Native people. *Pharmacogenet. Genomics* **22**, 429–440 (2012).
89. Wassenaar, C.A., Zhou, Q. & Tyndale, R.F. CYP2A6 genotyping methods and strategies using real-time and end point PCR platforms. *Pharmacogenomics* **17**, 147–162 (2016).
90. Application note – HiFi amplicon for pharmacogenetics: CYP2D6 <https://www.pacb.com/wp-content/uploads/Application-note-HiFi-amplicon-sequencing-for-pharmacogenetics-CYP2D6.pdf> Accessed January 8, 2024.
91. Haberl, M. *et al*. Three haplotypes associated with CYP2A6 phenotypes in Caucasians. *Pharmacogenet. Genomics* **15**, 609–624 (2005).
92. Twist Alliance Long-Read PGx Panel BED File. Twist Bioscience. Twist Bioscience <https://www.twistbioscience.com/resources/data-files/twist-alliance-long-read-pgx-panel-bed-file>.
93. Turner, A.J. *et al*. Characterization of complex structural variation in the. *Front. Pharmacol.* **14**, 1195778 (2023).
94. Twesigomwe, D. *et al*. StellarPGx: a Nextflow pipeline for calling star alleles in cytochrome P450 genes. *Clin. Pharmacol. Ther.* **110**, 741–749 (2021).
95. Hari, A. *et al*. An efficient genotyper and star-allele caller for pharmacogenomics. *Genome Res.* **33**, 61–70 (2023).
96. Lee, S.B., Shin, J.Y., Kwon, N.J., Kim, C. & Seo, J.S. ClinPharmSeq: a targeted sequencing panel for clinical pharmacogenetics implementation. *PLoS One* **17**, e0272129 (2022).
97. Patterson, M. *et al*. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. *J. Comput. Biol.* **22**, 498–509 (2015).

98. Mwenifumbo, J.C. *et al*. Novel and established CYP2A6 alleles impair in vivo nicotine metabolism in a population of Black African descent. *Hum. Mutat.* **29**, 679–688 (2008).
99. Piliguian, M. Identification and Characterization of Novel CYP2A6 Variants in African American Slow Nicotine Metabolizers. University of Toronto <https://tspace.library.utoronto.ca/handle/1807/44055> (2014).
100. Piliguian, M. *et al*. Novel CYP2A6 variants identified in African Americans are associated with slow nicotine metabolism in vitro and in vivo. *Pharmacogenet. Genomics* **24**, 118–128 (2014).