



Published in final edited form as:

*Clin Pharmacol Ther.* 2020 January ; 107(1): 154–170. doi:10.1002/cpt.1643.

## PharmVar GeneReview: *CYP2D6*

Charity Nofziger<sup>1,\*</sup>, Amy J. Turner<sup>2,3,\*</sup>, Katrin Sangkuhl<sup>4</sup>, Michelle Whirl-Carrillo<sup>4</sup>, José A. G. Agúndez<sup>5</sup>, John L. Black<sup>6</sup>, Henry M. Dunnenberger<sup>7</sup>, Gualberto Ruano<sup>8</sup>, Martin A. Kennedy<sup>9</sup>, Michael S. Phillips<sup>10</sup>, Houda Hachad<sup>11</sup>, Teri E. Klein<sup>4</sup>, Andrea Gaedigk<sup>12,#</sup>

<sup>1</sup>PharmGenetix GmbH, Nideralm, Austria <sup>2</sup>Department of Pediatrics, Section of Genomic Pediatrics and Children's Research Institute, The Medical College of Wisconsin, Milwaukee, Wisconsin, USA. <sup>3</sup>RPRD Diagnostics LLC, Wauwatosa, Wisconsin, USA. <sup>4</sup>Department of Biomedical Data Science, Stanford University, Stanford, CA, USA <sup>5</sup>University Institute of Molecular Pathology Biomarkers, UEx, Cáceres; ARADyAL Instituto de Salud Carlos III. Spain. <sup>6</sup>Personalized Genomics Laboratory, Division of Laboratory Genetics and Genomics, Mayo Clinic laboratories, Mayo Clinic, Rochester MN (200 1st Street SW, Rochester MN 55902) <sup>7</sup>Mark R. Neaman Center for Personalized Medicine, NorthShore University HealthSystem, Evanston, IL, USA <sup>8</sup>Institute of Living at Hartford Hospital, Genomas Laboratory of Personalized Health, Hartford, Connecticut (67 Jefferson Street, Hartford, Connecticut 06106) <sup>9</sup>Department of Pathology and Biomedical Science, University Otago, Christchurch, New Zealand <sup>10</sup>Sequence Bioinformatics Inc., 139 Water Street, 2nd Floor, St. John's NL, A1C 1B2, Canada <sup>11</sup>Translational Software, Bellevue, Washington <sup>12</sup>Division of Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy Kansas City, Kansas City and School of Medicine, University of Missouri-Kansas City, Kansas City, MO, USA

### Abstract

The Pharmacogene Variation Consortium (PharmVar) provides nomenclature for the highly polymorphic human *CYP2D6* gene locus. *CYP2D6* genetic variation impacts the metabolism of numerous drugs and thus can impact drug efficacy and safety. This GeneReview provides a

---

#corresponding author: [agaedigk@cmh.edu](mailto:agaedigk@cmh.edu).

\*co-lead authors

Supplemental Materials

(Report Descriptions.pdf)

Report Template Descriptions

(Genotype Reporting Template.xlsx)

*CYP2D6* Genotyping Method and Data Templates

(G2P Reporting Template.xlsx)

*CYP2D6* Genotype to Phenotype Translation Template

This template provides a standardized way of reporting how *CYP2D6* genotype was translated into phenotype (G2P, genotype to phenotype). Although CPIC-recommends translating genotype into phenotype using the consensus method jointly developed by the CPIC, the DPWG and a panel of experts (48), not every investigator is utilizing this method. It is imperative, however, that the method of how genotype is translated (i.e. which genotypes were classified into which phenotype group) is provided in detail. Please enter data into the first tab labeled '*CYP2D6* G2P translation\_template' (G2P, genotype to phenotype). The second tab labeled '*CYP2D6* G2P translation\_examples' provides a data set to exemplify how data may be presented. In this example, genotypes are sorted by their assigned Activity Score and respective phenotype assignments (UM, NM, IM, and PM) per CPIC recommendations. The template is also configured to provide genotype frequencies. The table can be revised as needed to accommodate the genotypes detected in a study.

This table can be published in a revised form in a manuscript methods section, or as supplemental table. Please remove the 'example tab' before publishing.

comprehensive overview and summary of *CYP2D6* genetic variation and describes how the information provided by PharmVar is utilized by the Pharmacogenomics Knowledgebase (PharmGKB) and the Clinical Pharmacogenetics Implementation Consortium (CPIC).

---

## CYP2D6 brief history

In the 1970s polymorphic oxidation of the antihypertensive agent debrisoquine and the antiarrhythmic agent sparteine was first described (1, 2). It was observed that a small percentage of individuals had ‘deficient’ metabolism of these drugs and were thus termed ‘poor metabolizers’ (PMs). Similar observations soon followed for numerous other drugs (3). Human CYP2D6 protein was eventually purified in 1984 (4) and the gene mapped to chromosome 22q13 in 1987 (5). Two years later, the *CYP2D6* gene was cloned and sequenced (6) and discovered that the gene locus contains two additional genes, the nonfunctional *CYP2D7* gene, and the *CYP2D8* pseudogene. Not long thereafter the first genetic variations responsible for the PM phenotype were identified (7). In 1990, the first test for detecting *CYP2D6* allelic variation (alleles now known as *CYP2D6*\*3 and \*4) was published (8). In the early 1990s, technical advancements, namely polymerase chain reaction (PCR) and Sanger sequencing, revealed additional variants in the *CYP2D6* gene in subjects presenting with no or diminished metabolic capacity. Anticipating an ‘explosive number’ of variants to be discovered, a group of experts recognized the need to systematically catalog the growing number of variants which led to the 1996 landmark paper by Daly and colleagues describing a system to track *CYP2D6* allelic variation, which became known as the star (\*) allele nomenclature system (9). This haplotype-based system was readily embraced by the field and became the standard for all *CYP* genes (9). The Human Cytochrome P450 nomenclature webpage was launched in 2001 (10) and served the global pharmacogenetics (PGx) community until it was transitioned to PharmVar in 2017 (11).

## Status of nomenclature before PharmVar

In its 15 year period of operation, over one hundred unique star alleles (not counting suballeles) were cataloged for *CYP2D6* by the Human Cytochrome P450 nomenclature webpage. Initially, only exons were required to be sequenced. As sequencing the entire gene became more routine, single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) (collectively referred to as single/small nucleotide variants, or SNVs, from here onward) located in the upstream and intronic gene regions were also submitted and included in haplotype definitions. It became difficult, however, to know whether an allele definition without SNV annotations in these regions did not in fact have any SNVs, or whether these regions were not sequenced when the allele was first defined. Also, the P450 nomenclature webpage eventually ceased to list suballeles, upon which the community missed out on information that may be important for assay design and/or the interpretation of sequence data. The webpage, however, was missing information and features that were considered to be ‘must-haves’ (11). Furthermore, the need to establish consistent allele definition criteria that could be applied across pharmacogenes was recognized. Still, the information provided by the P450 nomenclature webpage was highly valuable and heavily

relied upon by knowledge resources like PharmGKB (12) as well as the pharmacogenetic testing and implementation communities (e.g. reporting of clinical test results).

## Clinical relevance

There are many cytochrome P450 enzymes encoded in the human genome that carry out a wide range of oxidative metabolic processes, including the biotransformation of endogenous molecules, dietary components and drugs or prodrugs (13–16). Of these, CYP2D6 is among the most extensively studied and is arguably one of the most important drug metabolizing enzymes (17). CYP2D6 activity can range from complete absence to increased activity (18–20) and this marked inter-individual variation in activity can have significant clinical consequences (13, 21).

CYP2D6 contributes to the metabolism of many drugs including antidepressants (e.g., paroxetine, fluoxetine, venlafaxine), a number of atypical and typical antipsychotics (e.g., aripiprazole, clozapine, pimozide), antineoplastic agents (e.g., tamoxifen), adrenergic antagonists (e.g., carvedilol, metoprolol), and analgesics (e.g., codeine and tramadol). We refer to the PharmGKB drug label annotations (22), the FDA table on Pharmacogenomic Biomarkers in Drug Labeling (23), and the CPIC drug-gene pairs pages (24) for more information. Links to resources mentioned here and throughout this GeneReview are summarized in Table 1.

## Drug-drug interaction, polypharmacy, and enzyme inhibition and induction

A patient's metabolic profile may be profoundly impacted by one or multiple co-medications. In psychiatry, for example, there is widespread combined treatment with both antipsychotics and antidepressants. If both (or multiple) drugs are CYP2D6-dependent, each may be metabolized at a lower rate. CYP2D6 metabolic capacity may also be compromised by CYP2D6-inhibiting drugs. PMs are generally least affected by drug-drug interactions and drug inhibitors of the polymorphic enzyme, as these have no function to inhibit. In contrast, ultrarapid (UMs), normal (NMs) and intermediate metabolizers (IMs) often pheno-convert to IM or PM status (25–28). Fluoxetine, for example, is not only metabolized by CYP2D6, but also acts as a non-reversible inhibitor that causes most subjects to convert to PM status (29, 30). For more information on CYP2D6 inhibitors we refer to the Drug Interactions Flockhart Table™ (31). Some common dietary supplements including herbal remedies may also become clinically significant by inhibiting CYP2D6 activity, e.g. sesamin, turmeric, lotus herbals in cosmetics and teas (32, 33). While many agents have been known for their inducing capabilities towards many Cytochrome P450 enzymes, there are no known clinical CYP2D6 inducers.

## CYP2D6 and the Clinical Pharmacogenetics Implementation Consortium (CPIC)

CPIC develops detailed, evidence-based clinical practice guidelines for drugs affected by genetic variation (34, 35). Numerous *CYP2D6* gene-drug pairs have been prioritized through consideration of multiple factors, such as the available body of PGx knowledge, severity of

the clinical consequences, availability of alternative therapies or whether a prescribing change (drug choice or dose) is warranted. To date, six guidelines have been published for *CYP2D6* covering the following drugs: selective serotonin reuptake inhibitors (SSRIs) (fluvoxamine, paroxetine, and sertraline) (36), tricyclic antidepressants (TCAs) (amitriptyline, clomipramine, desipramine, doxepin, imipramine, nortriptyline and trimipramine) (37), atomoxetine (38), codeine (39), tamoxifen (40) and ondansetron and tropisetron (41). All guidelines and supplemental materials are freely available on the CPIC website (42). Each guideline has multiple components with *CYP2D6* phenotype-specific therapeutic recommendations at its core, access to the reviewed evidence, and implementation resources to support the translation of the guideline into electronic health records (EHRs), as well as example clinical decision support text. *CYP2D6*-drug pairs of interest for future guidelines are also listed on the CPIC website and include antipsychotics, additional antidepressants and opioids, and beta-blockers.

## CYP2D6 and the Pharmacogenomics Knowledgebase (PharmGKB)

PharmGKB collects, curates and disseminates knowledge about the impact of human genetic variation on drug response (12). The PharmGKB *CYP2D6*-dedicated webpage allows structured access to gene-specific PGx knowledge (43). Information is presented in sections including prescribing information, drug label annotations, clinical annotations, variant annotations, and curated pathways. Prescribing information encompasses 1) annotations of clinical guidelines from sources such as CPIC, the Royal Dutch Association for the Advancement of Pharmacy - Pharmacogenetics Working Group (DPWG), and the Canadian Pharmacogenomic Network for Drug Safety (CPNDS) and 2) “Rx study annotations” that provide genotype-based drug dosing or prescribing information reported in individual journal articles. Fourteen CPIC, 19 DPWG, and 2 CPNDS clinical guideline annotations as well as 7 Rx study annotations, are available for *CYP2D6* with overlapping *CYP2D6*-drug pairs. PharmGKB extracts PGx-relevant information from agency-approved drug labels and applies a PGx-level tag (such as Testing Required, Testing Recommended, Actionable PGx, or Informative PGx) based on the interpretation of the level of action implied in each label. On the *CYP2D6* page, annotations can be accessed for 69 FDA approved labels, 19 EMA approved labels, 9 PMDA approved labels, and 16 HCSC approved labels (see (22) for more details and updated information). Currently, PharmGKB contains 92 *CYP2D6*-related clinical annotations, i.e. evidence-based level summaries for specific allele-drug combinations. Pharmacokinetic pathways, available for 34 drugs, highlight a major or minor contribution of *CYP2D6* to their metabolism. PharmGKB and CPIC work together to develop gene-specific resources that accompany each CPIC guideline, including allele definition mapping, allele functionality, allele frequency and diplotype to phenotype mapping files with a standardized format. Gene-specific information tables for *CYP2D6* are also available from PharmGKB.

## Genotype to phenotype translation

An individual has two *CYP2D6* haplotypes, one on each chromosome, which constitute his/her diplotype. For example, a *CYP2D6*\*2/\*4 diplotype assignment signals that one chromosome (or allele) carries SNVs defining the *CYP2D6*\*2 haplotype and the second

chromosome (or allele) carries SNVs defining the *CYP2D6\*4* haplotype. The term ‘genotype’ can refer either to the sum of all detected SNVs or to a person’s diplotype. However, the term ‘genotype’ is often used interchangeably with ‘diplotype’ when describing a person’s genetic status.

For functional classification, individuals are categorized into four phenotype groups, i.e. PMs, IMs, NMs (formerly extensive metabolizers) and UMs (44). To facilitate the translation process, a genotype can be converted into an Activity Score, which is then translated into phenotype. The AS system was first published in 2008 (45, 46) and widely adopted in the field including CPIC. Briefly, a value is assigned to each allele reflecting no function (0), decreased function (0.25; 0.5) and normal function (1). The sum of the values assigned to each allele is a genotype’s (or diplotype’s) AS. For example, a *CYP2D6\*1/\*5* genotype has values of 1 (*\*1*, normal function allele) and 0 (*\*5*, no function allele) giving rise to an AS of 1, while a *CYP2D6\*1×2/\*5* genotype has an AS of 2 due to doubling the value assigned to the *\*1×2* allele carrying a gene duplication. The Diplotype-Phenotype-Table provided by the PharmGKB and CPIC serves as a reference for calculating the AS of each genotype (47).

In the past, there was no consensus of how to translate a genotype or AS into one of the four phenotype (metabolizer status) categories. This holds especially true for subjects with AS=1 genotypes (i.e., heterozygous for one no function and one normal function allele, or two decreased function alleles), which were categorized as either NMs or IMs by various end-users (CPIC, DPWG, clinical laboratories, researchers, physicians, patients, etc.). Inconsistent phenotype categorization has caused confusion and may also have held back wide-spread *CYP2D6* PGx implementation efforts. To address this, a working group of international experts recently developed a consensus method for translating *CYP2D6* genotype into phenotype (48, 49). The working group has also assessed whether certain decreased function alleles should receive a lower value for AS calculation to more accurately reflect substantially decreased function. Consensus was reached to ‘downgrade’ the activity value for the *CYP2D6\*10* allele from 0.5 to 0.25 for AS calculation. Activity values for other decreased function alleles such as *CYP2D6\*41* will be reviewed in the future. The consensus method jointly recommended by CPIC and the DPWG translates genotype to phenotype as follows: UMs, AS > 2.25; NMs, AS 1.25 x 2.25; IMs, AS 0 < x < 1.25 and PMs; AS = 0. There is no rapid metabolizer phenotype group due to the lack of evidence supporting this metabolizer category (48, 49).

## Gender and age-related differences

Some studies suggest gender-related differences in *CYP2D6* expression (50), and it has been reported that *CYP2D6* activity increases during pregnancy (51). *CYP2D6* is not appreciably expressed during fetal development; however, activity is measurable at 2 weeks post-partum and reaches levels that are comparable to those of adults within the first year of life (52).

## Need for standardized genetic variation definitions and the reporting of their functional/clinical impacts

In order to guide drug therapy, it is imperative to understand the complex *CYP2D6* gene locus and allelic variation as well as allele and genotype function (18, 53). While many alleles have been observed in phenotypic PMs and their underpinning genetic variations described (e.g. *CYP2D6*\*3, \*4, \*5, etc.) the function of many allelic variants remains unknown or uncertain. Series of variants have been investigated *in-vitro*, but results can be inconsistent among test systems and substrates (54–56) (see *CYP2D6* functionality table for a detailed summary (47)). Furthermore, although *in-silico* prediction tools are improving (57, 58), *in-vivo* validation is still the gold standard. The determination of the effects of two or more copies of normal or decreased function and other rare alleles remains a challenge since they are under-represented or absent in most studies.

Another confounding variable is substrate-specificity, i.e. the activity of an allele may substantially differ between substrates. One example is *CYP2D6*\*10 which exhibits different levels of decreased function toward a range of substrates. This allele was noted for little activity towards tamoxifen triggering specific recommendations by CPIC for *CYP2D6*\*10-containing genotypes for tamoxifen (40). As shown by Hertz and colleagues, subjects with the *CYP2D6*\*2 and \*17 alleles also had reduced activity towards tamoxifen (59). Another example is the decreased function allele *CYP2D6*\*17, which has been reported to metabolize risperidone at a normal or even increased rate (15, 60). Therefore, for any given allele, caution should be taken when extrapolating functional data from one drug or substrate to another. In a perfect world, one would be able to assess the *in-vivo* function of each individual *CYP2D6* haplotype with each individual *CYP2D6* substrate. Such a possibility would exponentially refine the phenotype predicting capacity of *CYP2D6* genetic testing.

Furthermore, the impact of co-medications (drug-drug interactions) may also not affect all allelic variants equally, posing yet another knowledge gap. Finally, there is still limited or no information regarding genetic variability for many minority populations (61, 62).

A solid foundation of the genetic variations occurring within a gene encompasses not only the cataloging of the existing genetic variations, but also their precise arrangement into haplotypes, and that haplotype's impact on enzyme function. Current reality is that the discovery of *CYP2D6* genetic variants far exceeds that of determining their functional impact. The ability to accurately sequence genomic DNA in a massive and parallel manner combined with the lack of reliable and standardized functional assay methodology largely contributes to this imbalance.

Although clinical PGx programs have successfully been implemented over past years, numerous challenges remain to accelerate adoption (63). Standardization of various implementation areas represents an opportunity for all PGx stakeholders to expand and improve laboratory processes, test ordering, results reporting and data representation. In particular, the accurate representation of clinically actionable PGx information in the patient's health record calls for harmonization at the genotype and phenotype levels (64, 65). While many PGx laboratories utilize the nomenclature recommended by PharmVar, inter-

laboratory differences remain in testing approaches. Clinical PGx testing for *CYP2D6* is performed on a variety of platforms using different methodologies such as panel or array-based SNV genotyping and both Sanger and next-generation sequencing (NGS) applications. For more information regarding *CYP2D6* test platforms and selection see Bousman et al (66, 67). Genotyping data can be reported as chromosomal position or genomic position on a gene's reference sequence (RefSeq), amino acid change, rs ID, and/or most commonly for *CYP2D6*, as a haplotype using star nomenclature. There are currently no standards, however, which SNVs need to be tested to accurately identify haplotypes or which haplotypes should be tested at a minimum (which may vary among populations). Recommendations have been published by the Association for Molecular Pathology and College of American Pathologists for *CYP2C9* and *CYP2C19* testing (68, 69); recommendations for *CYP2D6* will be forthcoming in the future. The lack of testing standards can result in considerable variation in the regions and SNVs interrogated, which may lead to different genotype assignments and phenotype predictions. This is not only of concern for clinical testing but also the interpretation of research findings.

The star allele nomenclature is a simple, short-hand method for describing all SNVs in a given allele/haplotype and provides a system to communicate alleles tested and detected in reports and publications. Standardizing the nomenclature and requirements for allele designation for previously established, as well as new alleles, will help to ensure that each star allele represents a unique and fully defined haplotype. It will also help to minimize "mis-interpretation" of a genotype result and its clinical implication. Many PGx constituents are recognized consumers of *CYP2D6* allele definitions and they often depend on one another for achieving their scientific activities or creating their tools (70). Thus, using the widely accepted PharmVar nomenclature system consistently throughout the PGx space allows all stakeholders to 'speak the same language'.

Clinicians and patients are arguably the two end-user groups that benefit the most from standardized allele designations. Standardization will increase clinician and patient confidence in, and the understanding of, PGx tests by reducing confusion around the genetic variation(s) being reported. Consistent nomenclature is a prerequisite for comprehensive EHR integration and interoperability, as well as for the establishment of clinical decision support algorithms and the design of clinical support tools such as interruptive alerts (64, 65). The design of drug/allele combinations for alerts, in particular, will require detailed annotation of nomenclature into the EHR. Finally, as outcome data are accumulated into the EHR, analysis of PGx clinical correlations will require a harmonized nomenclature of the gene variants. Such efforts will increase the likelihood that a patient's genetic variation profile will be similarly interpreted by different clinicians as they move across healthcare systems.

## The *CYP2D6* gene locus

Although the *CYP2D6* gene is relatively small (4382 bp) and possesses only nine exons, genetic analysis of this highly polymorphic gene locus is not trivial. The presence of the highly similar *CYP2D7* and *CYP2D8* genes within the locus (6) (Figure 1A) requires particular attention to prevent their co-amplification during PCR that may cause erroneous

variant calls. A hallmark feature of *CYP2D7* is a T-insertion in exon 1 that renders the gene nonfunctional; other regions with noticeable differences compared to *CYP2D6* include intron 6, exon 9 and the downstream region. Although, a two-step protocol involving the generation of a *CYP2D6*-specific amplicon by long-range (XL)-PCR and subsequently genotype analysis of the XL-PCR product (71, 72), largely circumvent this issue, this approach is time-consuming and not amenable to high-throughput testing. Commercially available *CYP2D6*-specific TaqMan assays were eventually developed simplifying genotyping; however, this (and similar methods) require the amplification of relatively short PCR fragments, which makes *CYP2D6*-specific primer design difficult for regions that are virtually identical with *CYP2D7*. Furthermore, it is nearly impossible to avoid interference of TaqMan assays by one of the many, often rare SNVs within the gene (73–78). In addition to utilizing differences in exons or introns to discriminate *CYP2D6* from *CYP2D7*, there are also structural features in their downstream regions that can be exploited for gene-specific amplification including the 1.6 kb long ‘spacer’ sequence downstream of *CYP2D7* that is absent in *CYP2D6* (Figure 1A). Finally, it is also important to realize that a number of SNVs present in *CYP2D6* alleles correspond to what is considered the wild-type *CYP2D7* sequence, which further challenges *CYP2D6*-specific amplification (79).

The *CYP2D6* gene locus is also afflicted by a vast array of gene deletions, duplications and gene rearrangements, collectively referred to as structural variants which cause gene copy number variation (CNV). Deletion and duplication events (Figure 1B and C) including the entire gene deletion (*CYP2D6\*5*), duplication and multiple copies of entire genes (e.g. *CYP2D6\*1xN*, *\*2xN* and *\*4xN*), as well as gene copies consisting of portions of *CYP2D6* and *CYP2D7* (commonly referred to as *CYP2D6–2D7* and *CYP2D7–2D6* hybrids) (72, 80–89) have been reported. To complicate matters even more, hybrids can be found on their own (as single entities) or in combination with other gene copies (tandems) (Figure 1C). The reader is directed to the PharmVar ‘Structural Variation’ document (Table 1), which provides up-to-date information about *CYP2D6* CNVs, including graphical sketches and references. Since these structures cannot be accommodated by the PharmVar database display (i.e. listing all nucleotide differences compared to the *CYP2D6* RefSeq), deletion and hybrid genes are annotated as e.g. ‘*CYP2D6* full gene deletion’ (*CYP2D6\*5*) or ‘*CYP2D7–CYP2D6* hybrid genes’ (*CYP2D6\*13*).

## CYP2D6 allele, genotype and phenotype frequencies across populations

The *CYP2D6* frequency table available at PharmGKB (47), summarizes population-based allele frequencies reported in the literature. Studies were considered for inclusion if 1) the ethnicity of the population was clearly indicated, 2) either allele frequencies or genotype frequencies were reported, 3) the methodology by which the genes were genotyped was indicated, 4) the sample population consisted of at least 50 individuals with a few exceptions (smaller cohorts that were part of larger studies, for example) and (5) the study represented an original publication. The ethnicities/locations reported in the articles are mapped into seven geographically defined groups (American, Central/South Asian, East Asian, European, Near Eastern, Oceanian, and Sub-Saharan African) and two admixed groups (African American/Afro-Caribbean and Latino) using the biogeographical grouping system developed by PharmGKB (90). The *CYP2D6* frequency table is periodically updated and



contains multiple tabs summarizing ‘allele frequencies by biogeographical group’, ‘diplotype frequencies by biogeographical group’, ‘phenotype frequency’ and ‘references’; the latter describes allele frequencies for each publication included in the listing, which also allows the user to customize allele frequencies as needed. There are a number of limitations regarding the accuracy of allele frequencies as follows: 1) based on published allele frequency data (limited for some populations); 2) most studies test for a limited number of allelic variants, which leads to an over-estimation of ‘default’ allele assignments (e.g. g. 100C>T (rs1065852) occurs on numerous alleles including *CYP2D6\*10*, but unless additional SNVs are tested, an allele may be assigned as *CYP2D6\*10* by ‘default’ (Figure 2A); likewise, if no SNVs are found, *CYP2D6\*1* is assigned which inflates the frequency of this allele; 3) inadequate testing for CNVs and 4) errors translating SNV results into star alleles, etc. Consequently, certain alleles may be over or under-reported or not detected at all. Therefore, all calculations based on allele frequencies are estimates at best and should be used with caution.

Regardless, there is profound variation among the calculated frequencies for individual alleles in the biogeographical groups. The decreased function alleles *CYP2D6\*17* and *CYP2D6\*29*, for example, are more prominent in the Sub-Saharan African and African American/Afro-Caribbean groups compared to others, while the *CYP2D6\*10* decreased function allele is the most common variant allele in the East Asian group. The nonfunctional *CYP2D6\*4* allele has the highest frequency in the European group, while the *CYP2D6\*1xN* duplication/multiplication allele, leading to increased function, is most often observed in the Oceanian group. Some alleles (e.g. *CYP2D6\*2*, *\*4* or *\*5*) are found at variable frequencies in almost every population studied, while others have only been found in some populations (e.g. *CYP2D6\*44* and *\*49* in Asians) to date. Finally, populations including South Africans, Caribbean’s and others with diverse founding populations and admixture often reveal unique allele frequency patterns (91–93).

Genotype frequencies are the result of allele frequencies in a given population and can be calculated using the Hardy Weinberg equation. Considering *CYP2D6\*1* through *\*139*, thousands of allele combinations are possible and thus, the number of genotypes in a given population or patient cohort can be quite large, especially in racially admixed populations. However, the actual number of combinations that occur in a given population may be significantly less depending on the number of alleles and their frequencies. Phenotype frequencies across populations are provided in the ‘Calculated phenotype frequency’ tab in the PharmGKB/CPIC *CYP2D6* Frequency Table (47). We stress, however, that all phenotype group frequencies (including those shown in the PharmGKB/CPIC table) have to be viewed with caution due to the limitations regarding the accuracy of allele frequencies as well as the method used to translate genotype into phenotype and inconsistencies in the classification of ‘population’, ‘ethnicity’ or ‘race’ (94).

## Allele function

The PharmGKB/CPIC *CYP2D6* allele functionality table (47) captures the clinical function per allele together with supporting literature evidence. CPIC guideline development includes the process for determining clinical allele functionality based on published literature and

input from the guideline authors. The *CYP2D6* diplotype to phenotype table provides information to steer integration of the guideline content into clinical implementation systems. The table includes the mapping of each diplotype to its resulting activity score, *CYP2D6* metabolizer phenotype, and EHR priority result notation. It also contains consultation text examples per metabolizer/Activity Score combination.

A distant SNP (rs5758550) located 116kb downstream of the *CYP2D6* gene locus has been reported to impact *CYP2D6* function (95). Whether this so-called ‘enhancer’ SNP is clinically relevant, and warrants changes of allele function assignments remains to be seen.

## PharmVar nomenclature and *CYP2D6* allele designation

PharmVar uses a number of conventions for storing and displaying allelic data consistently across genes relying on public standards and data sources wherever possible (for additional information see the ‘standards’ document (96)). The standardized nomenclature follows criteria developed by gene experts. The ‘Allele Designation and Evidence Level Criteria’ document describes the nomenclature system and provides examples (97)). For instance, a new star number is only issued if a haplotype contains a SNV that 1) results in an amino acid change (e.g. *CYP2D6\*118* harbors two SNVs, one of which causes an amino acid change (T310A)); 2) abolishes a splice site (e.g. g.1847G>A (rs3892097) in intron 3 of *CYP2D6\*4* alters a splice site that leads to a frameshift and premature translation termination) or has been shown to alter function through alternative splicing (e.g. g.2989G>A (rs28371725) in intron 6 of *CYP2D6\*41*) or 3) changes expression levels causing decreased or increased function (no examples to date). In contrast, new haplotypes containing variants that obliterate function are catalogued under the same star number and catalogued as a suballele. For example, any allele carrying a novel SNV and g.1847G>A will be designated as a *CYP2D6\*4* suballele and considered nonfunctional regardless of the nature of the novel SNV.

## The PharmVar *CYP2D6* gene expert panel

Volunteers representing a wide range of interests including *CYP2D6* research, clinical testing and clinical implementation were recruited from the PharmVar membership. This diverse expert panel (98) is an integral part of the critical and transparent PharmVar review process (11). To facilitate standardized usage of allele nomenclature, the panel also includes experts representing PharmGKB and CPIC.

The *CYP2D6* expert panel was the first of its kind and held its inaugural teleconference in December of 2017 and met on a monthly basis thereafter. The initial tasks included the development of the standardized *CYP2D6* submission form and requirements of information needed for allele submission. The panel was also instrumental in the review and curation of information as it was transitioned from the P450 nomenclature legacy page (99) into the PharmVar database. Subsequently, the panel has reviewed over 25 submissions resulting in numerous new star alleles and suballele designations (Table 2). Specific efforts of the panel are described in more detail below.

## The PharmVar CYP2D6 gene page

PharmVar maps sequence variations for each gene to genomic and transcript reference sequences (RefSeqs) issued by the NCBI Reference Sequences database (100), the GRCh37 and GRCh38 genome builds and the M33388 legacy reference sequence. For *CYP2D6*, PharmVar uses NG\_008376.3 (corresponding to AY545216) as the RefSeq for genomic DNA and NM\_000106.5 for mRNA/cDNA. Of note, NG\_008376.3 deviates from the M33388 legacy RefSeq that was initially used by the P450 Nomenclature webpage for allele definition (original content from the cypalleles.ki/se site is available through the archive link on the PharmVar homepage). As described in detail in the gene information document ('Read Me for *CYP2D6*'), genetic variation coordinates are shifted up (plus) or down (minus) in different gene regions due to these deviations. For example, the functional variant defining the *CYP2D6*\*4 allele maps to g.1846G>A in M33388 and to g.1847G>A in NG\_008376.3. PharmVar strongly encourages the PGx community to utilize NG\_008376.3-based coordinates moving forward.

PharmVar has obtained a Locus Reference Genomic (LRG) record for *CYP2D6* from the LRG Project, an NCBI (RefSeq) and EMBL-EBI Ensembl/GENCODE (EMBL-EBI) initiative (101). LRGs are universally accepted reference standards that are created specifically for clinical reporting by manual curation. LRGs are stable entities that never change or version. The recently issued LRG for *CYP2D6* (LRG\_303) will be used by PharmVar as the 'gold-standard' reference sequence in the future. LRG\_303 matches the most recent RefSeq NG\_008376.4 and contains an additional 819 bp of upstream and 1540 bp of downstream sequence compared to the NG\_008376.3 RefSeq currently used by PharmVar.

In this gene summary report, SNV positions are provided according to their location on NG\_008376.3 with the ATG start codon being +1. On the gene page, the user can easily cross-reference position(s) by choosing the sequence or genome build of interest; there is also the option of two count modes, i.e. counting from the first nucleotide in the sequence or the ATG translation start codon being +1.

Of importance, the *CYP2D6* sequence in GRCh37 matches *CYP2D6*\*2.011 (formerly \*2M), which has numerous SNVs compared to the *CYP2D6*\*1 reference allele (NG\_008376.3). When reporting star alleles using GRCh37, these differences need to be accounted for. For example, g.2851C>T (rs16947) is present on *CYP2D6*\*2 and numerous other haplotypes and is not identified as a variant when compared to GRCh37.

Furthermore, insertions and deletions of nucleotides in a repeat or homopolymer sequence are consistently listed by PharmVar using the 3' rule that is recommended by the Human Genome Variation Society (HGVS). Briefly, an inserted or deleted base is always listed using the 3'-most position of the repeat sequence relative to the reference sequence (96). Previously, such insertions and deletions have been reported using either the 5' or the 3' alignment rule. Although either annotation generates the same translation effect, using different positions to report the same SNV is confusing as illustrated for the core SNVs characterizing *CYP2D6*\*9 and *CYP2D6*\*20 (Figure 3). Other examples are *CYP2D6*\*30

and \*58 which harbor a 9-base pair insertion and *CYP2D6*\*40 which possesses two copies of the 9-bp motif. Of note, annotations for indels may deviate between PharmVar and dbSNP due to different alignment modes as illustrated in Figure 3 (numbering is also further complicated due to the fact that the *CYP2D6* gene is encoded on the minus (–) strand). Insertions may also be annotated as duplications as is the case for the *CYP2D6*\*20G-insertion resulting in being reported as a G-duplication by dbSNP. Further information on alignment standardization is provided in the ‘Standards’ document (96).

The *CYP2D6* gene page (102) offers a wealth of information that is essential for understanding this complex gene locus including the ‘Read Me’, ‘Change Log’ and ‘Structural Variation’ documents which complement the information displayed in the database’s ‘Table View’. The following sections provide a comprehensive overview of ‘how PharmVar works’, highlight content from the aforementioned documents and exemplify particular challenges of *CYP2D6* nomenclature and standardized translation into phenotype.

## CYP2D6 haplotype evidence levels

The evidence level symbols displayed on the *CYP2D6* PharmVar gene page reflect all submissions PharmVar has received for that haplotype and indicate ‘Definitive’ (Def), ‘Moderate’ (Mod) and ‘Limited’ (Lim) levels of evidence in support for the definition of a given haplotype. This 3-category system represents a modified ClinVar classification system; more detailed information is provided in the ‘Allele Designation Criteria and Evidence Level’ document (97). This type of information, i.e. whether an allele was entirely sequenced and how the haplotype was determined, was not systematically captured prior to the creation of PharmVar. For existing haplotype definitions, an extensive literature review was conducted to assign evidence levels. Many alleles are currently labeled as ‘Lim’ because their definition was either 1) solely based on exon (including exon/intron boundaries) sequencing (this was the case for the majority of *CYP2D6*\*2 and \*4 suballeles, \*23-\*27, \*30, \*34, etc.), or 2) were entirely sequenced, but how the haplotype was determined was not described or unclear (e.g. *CYP2D6*\*52 and \*60). *CYP2D6*\*2.002 through \*2.009 show examples of the value of having these evidence levels. These *CYP2D6*\*2 suballeles, for instance, are shown as ‘Lim’ meaning that it is unknown whether these alleles indeed do not have any SNVs in upstream and/or intronic regions, which would starkly contrast from all other *CYP2D6*\*2 suballele definitions based on full-length sequence information. Furthermore, the allele known as *CYP2D6*\*2A (now designated \*2.001) is not the only suballele carrying g.–1584C>G (rs1080985). This SNV has also been identified on an allele that matched the *CYP2D6*\*2E definition (now designated \*2.005), as well as the newly designated \*2.012, \*2.013 and \*2.018 suballeles (Figure 4A). Another excellent example is *CYP2D6*\*32. A recent submission of a full-length sequence revealed that the initial allele designation (based on exon-only sequencing) may have missed an SNV in intron 6 that was subsequently reported in \*41 (g.2989G>A, rs28371725) and is presumed to interfere with splicing and causes decreased activity (Figure 4B). It remains unknown, however, whether the E410K amino acid change on *CYP2D6*\*32 (g.3854G>A, rs769157652) further decreases activity, compensates for the g.2989G>A splice defect, or has no impact. With this new knowledge, alleles previously called *CYP2D6*\*41 based on the presence of g.2989G>A could conceivably be *CYP2D6*\*32 unless the test also interrogated g.3854G>A, which

distinguishes the two alleles. These examples highlight challenges of defining alleles based on limited information, but also demonstrate the value of evidence levels, i.e. signaling to users that haplotypes labeled as ‘Lim’ and ‘Mod’ are to be viewed with caution.

PharmVar solicits submissions for all alleles labeled ‘Lim’ and ‘Mod’ to ultimately raise their evidence levels to ‘Def’. PharmVar also encourages encore submissions for ‘Def’ with single citations to further corroborate the allele’s definition.

## PharmVar IDs

Each previously cataloged *CYP2D6* haplotype received a PharmVar ID (PVID). The PVID is a unique numeric identifier similar to a dbSNP rs ID and can be thought of as an allele’s postal code. Because star allele names are driven by functional grouping, and since experimental evidence is not required for allele indexing, star allele names can be subject to change. A case in point is the nonfunctional *CYP2D6\*14A* allele. It was reclassified as *\*114.001* (PV00140) (103) because its function, only determined after it was originally named, differs from that of *\*14B* (*\*14.001*; PV00138; decreased function). Future star allele name changes may be necessary as functional data become available. If an allele’s star designation is updated to a new star number, the PVID of the haplotype remains constant. In contrast, a new PVID will be assigned if a haplotype definition changes, e.g. through the addition or removal of SNVs. An example is the aforementioned *CYP2D6\*32* allele. Its initial PVID (PV00141) was retired after its haplotype was updated, and a new PVID (PVID00456) subsequently assigned. Original PVIDs and their haplotype definitions can be tracked in the database via the PVID search function.

## Curation efforts

Extensive curation efforts were part of the content transfer from the P450 nomenclature webpage into the PharmVar database to standardize the annotations to the above-mentioned conventions (Table 3). The following sections describe general and specific efforts undertaken.

### Gene region mapped/required for allele definition:

Of concern was that some alleles were sequenced entirely while others, mostly those submitted prior to 2010, had only exon and intron/exon boundaries covered. To address this issue, the gene expert panel recommended to include 1600 bp of upstream and 250 bp of downstream regions as well as all exons and introns based on the following considerations 1) the requested 6068 bp long region (g.–1600 through g.4468) can easily be sequenced with today’s methods (for more information, see section ‘Methods for *CYP2D6* allele characterization’); 2) intronic SNVs impacting function have been discovered and may also be found on novel haplotypes (e.g. *CYP2D6\*32*); 3) SNVs in intronic, upstream or downstream regions may be utilized for allele discrimination and lastly 4) although it remains uncertain whether g.–1584C>G impacts function, its incorporation in many test panels warrants continued inclusion.

### Update to current RefSeq:

A second challenge concerned the upgrade to the then current RefSeq NG\_008376.3. As explained above, replacing M33388 which was historically used for mapping with NG\_008376.3 introduces shifts in SNV positions. To facilitate a smooth transition and easy cross-referencing to published M33388-based positions, PharmVar has developed multiple display features. For example, one can easily view the coordinates of a given genetic variant in relation to either the NG\_008376.3 or legacy M33388 RefSeqs with the click of a mouse. The chromosomal coordinates are also easily viewable for both GRCh37 and GRCh38, as well as the transcript RefSeq. In addition, coordinates for all RefSeqs are also viewable from both the sequence start and ATG start.

### Corrections, revisions, new alleles and other updates:

During the process of transitioning *CYP2D6* into the PharmVar database, comments and footnotes were removed, a number of errors were identified and corrected, and some alleles were revised based on additional published information that was either inadvertently omitted when first submitted to the CYP nomenclature webpage or never posted. In some cases, information published at a later time by other investigators was considered to further support an allele definition. References in support of allele definitions have been updated and those solely describing function removed (references for function are provided in the PharmGKB/CPIC *CYP2D6* Allele Functionality table (47)). A number of descriptors such as '*CYP2D7* intron 1' or '*CYP2D7* exon 9 conversion' have been replaced with the respective SNVs and all positions are now consistently mapped to NG\_008376.3 and aligned following the 3' rule. Changes and revisions are recorded in the 'Change Log' document (102).

Allele definitions based on partial sequence information will be replaced over time with full-length sequences. The 'Change Log' document tracks all "full-length" submissions and indicates the star alleles being replaced. Currently, more than 30 alleles have been updated by full-length sequence information (Table 3). The expert panel also recommended to reassign *CYP2D6*\*14A as *CYP2D6*\*114.001 because the function of this allele (no function) is different than the more common decreased function \*14B (\*14.001) allele.

As of August 2019, the *CYP2D6* expert panel has designated 25 novel alleles (\*115 through \*139) and 79 new suballeles. In addition, 33 alleles that were based on partial sequence information are now supported by completely sequenced alleles and their evidence level was raised from 'Lim' or 'Mod' to 'Def' (Table 3).

### Core allele definitions

For many alleles, there are a growing number of so-called suballeles that share one or more 'key' defining SNV, referred to from here on as 'core' SNVs. Suballele information can be valuable for the design of SNV genotyping assays and test platforms (sequence or genotype-based alike) as well as interpretation of test results. There is, however, no need to distinguish between suballeles for phenotype inference because all alleles under a star number are assumed to be functionally equal and therefore receive the same value for AS calculation.

Thus, even if a test is capable of distinguishing suballeles, they are generally reported as e.g. *CYP2D6\*2* instead of *CYP2D6\*2.002*, *CYP2D6\*2.003*, *CYP2D6\*2.004*, etc.

PharmVar and PharmGKB have collaboratively developed core allele definitions for each *CYP2D6* star allele. Only SNVs that 1) change an amino acid, impact function by changing expression levels or interfere with splicing and 2) are present in *all* suballeles within a star number group, are part of the core allele definition (Figure 4). With this rule-based system, suballeles can be collapsed into a single ‘core’ definition representing all suballeles categorized under a star number. For example, the growing number of *CYP2D6\*2* suballeles (currently 20) share two SNVs that fulfill the rules, i.e. g.2851C>T (R296C, rs16947) and g.4181G>C (S486T, rs1135840), and thus constitute the *CYP2D6\*2* core allele definition. For *CYP2D6\*4*, only g.1847G>A (causing aberrant splicing) is shared among the 28 subvariants that have been defined to date, and is, therefore, the sole variant of the *\*4* core allele definition.

Of importance, a sequence variant found in a core allele definition is not necessarily unique to that haplotype as illustrated by the two SNVs of the *CYP2D6\*2* core definition. Indeed, both SNVs are part of many other core allele definitions including, but not limited to, *CYP2D6\*8*, *\*11*, *\*17*, *\*29* and *\*41* as well as a number of suballeles. For example, g.2851C>T has been found on only one *CYP2D6\*4* suballele to date, *CYP2D6\*4.010*, while 4181G>C is present on the majority of *\*4* suballeles.

One challenge with core allele definitions is that some may change over time as new information becomes available. For instance, based on the rules described above, the current core allele definition for *CYP2D6\*11* contains g.882G>C (rs201377835), g.2851C>T (R296C) and g.4181G>C (S486T). According to the *CYP2D6* allele designation criteria, all new alleles carrying the detrimental g.882G>C splicing defect will be assigned as *\*11* suballeles, regardless of the nature of the other SNV(s) present. Consequently, the *CYP2D6\*11* core allele definition will change, for example, if a sequence that contains g.882G>C without g.2851C>T and/or g.4181G>C is identified. Another example is *CYP2D6\*32*, which was initially defined by exon sequencing only. The full-length sequence recently submitted to PharmVar revealed that this allele, in addition to its other core SNVs, also contains the functionally relevant g.2989G>A (rs28371725) in intron 6; this SNV was added to the *CYP2D6\*32* core allele definition (Figure 4). Of note, g.2989G>A is the core SNV of the decreased function *CYP2D6\*41* allele and is also part of other core allele definitions (i.e., *\*32*, *\*69*, *\*91*, *\*119*, *\*123*, *\*132* and *\*138*). These examples highlight how core allele definitions can be affected by new sequence submissions.

The core alleles are the basis of the *CYP2D6* allele definition table used in CPIC guidelines and by PharmGKB (this table is available to PharmGKB users as definition material). The *CYP2D6* core allele definitions are also utilized for clinical annotations in PharmGKB.

## The PharmVar Comparative Allele ViewEr

PharmVar has developed the Comparative Allele ViewEr (CAVE) tool to easily compare core alleles. This tool can be accessed using the “Compare View” button on the main

PharmVar page for each gene. The graphical display visualizes SNVs of interest contained in the core alleles. One prime example is g.100C>T which is part of numerous core allele definitions including *CYP2D6\*10*, *\*36* and *\*49*. Although the vast majority of *CYP2D6\*4* alleles carry g.100C>T, this SNV is not part of its core allele definition because one suballele, *CYP2D6\*4.012* (previously known as *CYP2D6\*4M*) lacks this SNV. The CAVE tool highlights core SNVs in blue color indicating their presence in respective core allele definitions (e.g. g.100C>T for *CYP2D6\*10*, *\*36* and *\*49*); those highlighted in gray color indicate that the SNV is present on one or more suballeles (e.g. g.100C>T for *CYP2D6\*4*).

In the graphical view mode, the user can choose any number of alleles for comparison via the CAVE selection pad. The graphical display also denotes whether a core SNV is known to alter function. More information and examples are provided in the ‘*CYP2D6* Read Me’ document (102) and in Figure 4C.

## Reporting genotype and translation into phenotype

In another collaborative effort, PharmVar and PharmGKB have developed templates to facilitate more consistent and transparent reporting of genotype details and how genotype is translated into phenotype (this information can be provided as supplemental materials of a publication to facilitate access to important data for subsequent curation). The first template file (Suppl materials 1) collects information including methods or platforms used for genotyping and which SNVs and CNVs are interrogated; the template also provides a standardized set-up for reporting genotype results for individual subjects, as well as allele frequencies. The second template file (Suppl materials 2) facilitates the reporting of how genotype is translated into an Activity Score and/or phenotype as well as genotype frequencies. Although CPIC and other groups champion the use of their standardized method (48), not every investigator or laboratory employs the CPIC recommended method. Too often, papers reference previous work stating that ‘genotyping was performed as previously described’ or indicate that ‘*CYP2D6* phenotype was correlated with the metabolism of a drug’ without specifying which SNVs or alleles were genotyped or how phenotype was assigned. The lack of such information makes it extremely difficult, if not impossible, to compare results with other published literature, or extract the necessary information for CPIC guideline development. Colleagues are therefore strongly encouraged to utilize the provided templates, or revised versions thereof, for publication of these types of information as supplemental materials.

## CYP2D6 reference materials

As pharmacogenetic testing becomes more routine, there is an increasing need for an established set of well-characterized reference materials for assay development, validation, quality control and proficiency testing. The Genetic Testing Reference Materials Coordination Program (GeT-RM), is a collaborative effort between the Centers for Disease Control and Prevention–based Genetic Testing Reference Material Coordination Program, Coriell Institute for Medical Research and members of the PGx testing community. A set of 137 genomic DNA samples (104) have previously been genotyped across several testing platforms, establishing a “consensus” genotype for 28 PGx relevant genes, including



*CYP2D6*. In a follow-up study focusing specifically on *CYP2D6*, inconclusive genotype calls and CNVs were resolved, along with 42 additional samples characterized. The latter contains rare *CYP2D6* alleles not previously represented in the panel, as well as a number of complex structural arrangements (105). Testing and research laboratories can acquire these materials from the Coriell Institute (Camden, NJ, USA).

## Inferring *CYP2D6* haplotype from Next Generation Sequence data and public databases

Bioinformatic tools have been developed to facilitate diplotype calling from Whole Genome Sequence (WGS) and/or NGS-based gene panel sequence data (e.g. Astrolabe (106), Stargazer (107), VCF Translator (108), Aldy (109) and Cypiripi (110)). These tools use computational approaches to infer the most likely diplotype based on the catalog of known haplotypes such as those defined by PharmVar. However, these algorithms have not been systematically evaluated and each faces its own set of limitations. High-quality NGS data and allele definitions are required in order to obtain accurate genotype calls. As reported by Cohn *et al.* (111), call rates from WGS for *CYP2D6* were lower compared to those of other genes which was mostly attributed to lower read depth and variant calling difficulties especially in the presence of CNVs and/or the presence of the *CYP2D6\*4* haplotype. Misalignment of *CYP2D7* or CNV reads and discarded reads due to multi-locus alignment are some of the major culprits, which may eventually be overcome with ever increasing read-lengths, improved NGS alignment tools and increasingly sophisticated diplotype calling algorithms along with a growing catalog of well-defined allele definitions. Given these particular challenges, haplotype calls using 1000 Genomes Project data (112) (and other data resources) should be interpreted with caution and are strongly suggested to be experimentally validated. One example is Coriell DNA NA19317 which was genotyped as homozygous for the *CYP2D6\*5* gene deletion by multiple platforms as part of a recent *CYP2D6* Get-RM project (105) but may be called as *CYP2D6\*2/\*2* when using vcf or bam files that have been generated from NGS data without a structural variant caller (unpublished observation). Furthermore, inferior WGS data, read alignment issues (e.g. misaligned *CYP2D7*), and possibly also *CYP2D6* regions that have been described as “inaccessible” (112) may contribute to erroneous diplotype calls or may trigger no-calls.

## Methods for *CYP2D6* allele characterization

While we abstain from reviewing *CYP2D6* genotyping techniques and platforms (see Bousman *et al.* (66, 67)), we offer a brief summary of approaches that have successfully been utilized to fully characterize novel allelic variants for PharmVar submission or confirm existing haplotype definitions. Although bioinformatic tools can be used to infer haplotypes, experimental validation is the gold standard for allele definition (see Allele Designation and Evidence Level document (97)).

As illustrated in Figure 5A and B haplotype is unequivocal for samples in which all SNVs, or all but one (not shown), are homozygous (or hemizygous in the presence of the *CYP2D6\*5* gene deletion). For samples with two or more heterozygous SNVs, computational phasing may not resolve haplotype and thus needs to be experimentally

determined. One validated approach is sequence analysis of long-range allele-specific PCR products. Briefly, allele-specific primers are employed to amplify one or the other allele (77, 113). Allele-specific amplicons can also be generated from duplicated gene copies and *CYP2D6-2D7* and *CYP2D7-2D6* hybrid genes (see examples provided by the *CYP2D6* Get-RM project (105)). All SNVs found on this PCR product represent that allele's haplotype (Figure 5C). Sequencing technologies such as single molecule real-time (SMRT) sequencing (114, 115) or Nanopore sequencing (116) have also been utilized to characterize *CYP2D6* haplotypes. Finally, haplotype may also be inferred by pedigree information. Figure 5C shows how inheritance can unequivocally determine which SNVs are located on an allele of interest. Finally, haplotype may also be established using long-range phasing of SNVs detected by WGS in conjunction with e.g. 10X Genomics long-distance phasing (117) across large haplotype blocks.

## Conclusions

This is the first of a series of gene-centric review articles focusing on important pharmacogenes. This inaugural summary provides essential information for the understanding of *CYP2D6* and its complex *CYP2D6* gene locus complementing the information provided by CPIC guidelines. We are highlighting PharmVar efforts of systematically cataloging *CYP2D6* allelic variation as well as collaborative efforts with the PharmGKB to make the information useful and easily accessible to the entire pharmacogenetics community.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

All authors have served on the PharmVar *CYP2D6* expert panel. We also thank Mara Hutz, PhD, for having served on the panel. In addition, we would like to sincerely thank Roger Gaedigk, PhD, for his assistance with graphical artwork.

### Funding

This work was funded by the National Institutes of Health for the Pharmacogene Variation Consortium (R24GM123930; PI, A.G.) and PharmGKB (R24GM61374; PI, T.E.K.).

### Conflicts of Interest:

C.N. is employed by Pharmgenetix Gmbh, an independent laboratory offering pharmacogenetic testing and reporting services. A.J.T.'s efforts are supported in part by RPRD Diagnostics, an independent clinical laboratory offering pharmacogenetic testing services. J.L.B. has intellectual property for AssureX Health (Myriad) and OneOme LLC and receives royalties for AssureX Health (Myriad). H.M.D. is a paid consultant for Admera Health and Veritas Genetics. G.R. is the founder and Medical Director of the Genomas Laboratory of Personalized Health. M.S.P. is an employee of Sequence Bioinformatics Inc. (St. John's NL, Canada) and has stock in the company. He also has stock in Genomic Medicine Ireland Ltd (Dublin, Ireland). H.H. is a paid employee and stockholder in Translational Software, a pharmacogenomics interpretive service. All other authors declared no competing interests for this work.

## References

- (1). Mahgoub A, Idle JR, Dring LG, Lancaster R & Smith RL Polymorphic hydroxylation of Debrisoquine in man. *Lancet* 2, 584–6 (1977). [PubMed: 71400]
- (2). Eichelbaum M, Spannbrucker N, Steincke B & Dengler HJ Defective N-oxidation of sparteine in man: a new pharmacogenetic defect. *Eur J Clin Pharmacol* 16, 183–7 (1979). [PubMed: 499318]
- (3). Brosen K & Gram LF Clinical significance of the sparteine/debrisoquine oxidation polymorphism. *Eur J Clin Pharmacol* 36, 537–47 (1989). [PubMed: 2570698]
- (4). Distlerath LM & Guengerich FP Characterization of a human liver cytochrome P-450 involved in the oxidation of debrisoquine and other drugs by using antibodies raised to the analogous rat enzyme. *Proc Natl Acad Sci U S A* 81, 7348–52 (1984). [PubMed: 6594694]
- (5). Eichelbaum M et al. Chromosomal assignment of human cytochrome P-450 (debrisoquine/sparteine type) to chromosome 22. *Br J Clin Pharmacol* 23, 455–8 (1987). [PubMed: 3472585]
- (6). Kimura S, Umeno M, Skoda RC, Meyer UA & Gonzalez FJ The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *Am J Hum Genet* 45, 889–904 (1989). [PubMed: 2574001]
- (7). Skoda RC, Gonzalez FJ, Demierre A & Meyer UA Two mutant alleles of the human cytochrome P-450db1 gene (P450C2D1) associated with genetically deficient metabolism of debrisoquine and other drugs. *Proc Natl Acad Sci U S A* 85, 5240–3 (1988). [PubMed: 2899325]
- (8). Heim M & Meyer UA Genotyping of poor metabolisers of debrisoquine by allele-specific PCR amplification. *Lancet* 336, 529–32 (1990). [PubMed: 1975039]
- (9). Ingelman-Sundberg M, Daly AK, Oscarson M & Nebert DW Human cytochrome P450 (CYP) genes: recommendations for the nomenclature of alleles. *Pharmacogenetics* 10, 91–3 (2000). [PubMed: 10739177]
- (10). Ingelman-Sundberg M, Oscarson M, Daly AK, Garte S & Nebert DW Human cytochrome P-450 (CYP) genes: a web page for the nomenclature of alleles. *Cancer Epidemiol Biomarkers Prev* 10, 1307–8 (2001). [PubMed: 11751452]
- (11). Gaedigk A et al. The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clin Pharm Ther* 103, 399–401 (2018).
- (12). Barbarino JM, Whirl-Carrillo M, Altman RB & Klein TE PharmGKB: A worldwide resource for pharmacogenomic information. *Wiley Interdiscip Rev Syst Biol Med* 10, e1417 (2018). [PubMed: 29474005]
- (13). Nebert DW, Wikvall K & Miller WL Human cytochromes P450 in health and disease. *Philos Trans R Soc Lond B Biol Sci* 368, 20120431 (2013). [PubMed: 23297354]
- (14). Saravanakumar A, Sadighi A, Ryu R & Akhlaghi F Physicochemical Properties, Biotransformation, and Transport Pathways of Established and Newly Approved Medications: A Systematic Review of the Top 200 Most Prescribed Drugs vs. the FDA-Approved Drugs Between 2005 and 2016. *Clin Pharmacokinet*, Epub 4 10 (2019). doi: 10.1007/s40262-019-00750-8
- (15). Zhou SF Polymorphism of human cytochrome P450 2D6 and its clinical significance: Part I. *Clin Pharmacokinet* 48, 689–723 (2009). [PubMed: 19817501]
- (16). Zhou SF Polymorphism of human cytochrome P450 2D6 and its clinical significance: part II. *Clin Pharmacokinet* 48, 761–804 (2009). [PubMed: 19902987]
- (17). Owen RP, Sangkuhl K, Klein TE & Altman RB Cytochrome P450 2D6. *Pharmacogenet Genomics* 19, 559–62 (2009). [PubMed: 19512959]
- (18). Hicks JK, Swen JJ & Gaedigk A Challenges in CYP2D6 Phenotype Assignment from Genotype Data: A Critical Assessment and Call for Standardization. *Curr Drug Metab* 15, 218–23 (2014). [PubMed: 24524666]
- (19). Teh LK & Bertilsson L Pharmacogenomics of CYP2D6: molecular genetics, interethnic differences and clinical importance. *Drug Metab Pharmacokinet* 27, 55–67 (2012). [PubMed: 22185816]
- (20). Zanger UM, Raimundo S & Eichelbaum M Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn-Schmiedeberg's Arch Pharmacol* 369, 23–37 (2004). [PubMed: 14618296]

- (21). Zhou ZW et al. Clinical association between pharmacogenomics and adverse drug reactions. *Drugs* 75, 589–631 (2015). [PubMed: 25895462]
- (22). PharmGKB CYP2D6 Drug Label Annotations. <https://www.pharmgkb.org/gene/PA128/labelAnnotation>
- (23). Pharmacogenomic Biomarkers in Drug Labeling. <https://www.fda.gov/drugs/science-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling>
- (24). CPIC Gene/drug pairs. <https://cpicpgx.org/genes-drugs/>
- (25). Gardiner SJ & Begg EJ Pharmacogenetics, drug-metabolizing enzymes, and clinical practice. *Pharmacol Rev* 58, 521–90 (2006). [PubMed: 16968950]
- (26). Zourkova A & Hadasova E Paroxetine-induced conversion of cytochrome P450 2D6 phenotype and occurrence of adverse effects. *Gen Physiol Biophys* 22, 103–13 (2003). [PubMed: 12870705]
- (27). Carrillo JA et al. Pharmacokinetic interaction of fluvoxamine and thioridazine in schizophrenic patients. *J Clin Psychopharmacol* 19, 494–9 (1999). [PubMed: 10587283]
- (28). Shah RR & Smith RL Inflammation-induced phenocconversion of polymorphic drug metabolizing enzymes: hypothesis with implications for personalized medicine. *Drug Metab Dispos* 43, 400–10 (2015). [PubMed: 25519488]
- (29). Brynne N, Svanstrom C, Aberg-Wistedt A, Hallen B & Bertilsson L Fluoxetine inhibits the metabolism of tolterodine—pharmacokinetic implications and proposed clinical relevance. *Br J Clin Pharmacol* 48, 553–63 (1999). [PubMed: 10583026]
- (30). Hara Y, Nakajima M, Miyamoto KI & Yokoi T Inhibitory effects of psychotropic drugs on mexiletine metabolism in human liver microsomes: prediction of in vivo drug interactions. *Xenobiotica* 35, 549–60 (2005). [PubMed: 16192107]
- (31). Drug Interactions Flockhart Table™. <https://drug-interactions.medicine.iu.edu/Main-Table.aspx>
- (32). Sasaki T, Sato Y, Kumagai T, Yoshinari K & Nagata K Effect of health foods on cytochrome P450-mediated drug metabolism. *J Pharm Health Care Sci* 3, 14 (2017). [PubMed: 28496987]
- (33). Ye LH et al. Identification and characterization of potent CYP2D6 inhibitors in lotus leaves. *J Ethnopharmacol* 153, 190–6 (2014). [PubMed: 24561383]
- (34). Caudle KE et al. Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process. *Curr Drug Metab* 15, 209–17 (2014). [PubMed: 24479687]
- (35). Relling MV & Klein TE CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther* 89, 464–7 (2011). [PubMed: 21270786]
- (36). Hicks JK et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. *Clin Pharmacol Ther* 98, 127–34 (2015). [PubMed: 25974703]
- (37). Hicks JK et al. Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants: 2016 update. *Clin Pharmacol Ther* 102, 37–44 (2017). [PubMed: 27997040]
- (38). Brown JT et al. Clinical Pharmacogenetics Implementation Consortium Guideline for Cytochrome P450 (CYP)2D6 Genotype and Atomoxetine Therapy. *Clin Pharmacol Ther* 106, 94–102 (2019). [PubMed: 30801677]
- (39). Crews KR et al. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450 2D6 genotype and codeine therapy: 2014 update. *Clin Pharmacol Ther* 95, 376–82 (2014). [PubMed: 24458010]
- (40). Goetz MP et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and Tamoxifen Therapy. *Clin Pharmacol Ther* 103, 770–7 (2018). [PubMed: 29385237]
- (41). Bell GC et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for CYP2D6 genotype and use of ondansetron and tropisetron. *Clin Pharmacol Ther* 102, 213–8 (2017). [PubMed: 28002639]
- (42). CPIC Guidelines. <https://cpicpgx.org/guidelines/>
- (43). PharmGKB CYP2D6 gene page. <https://www.pharmgkb.org/gene/PA128>

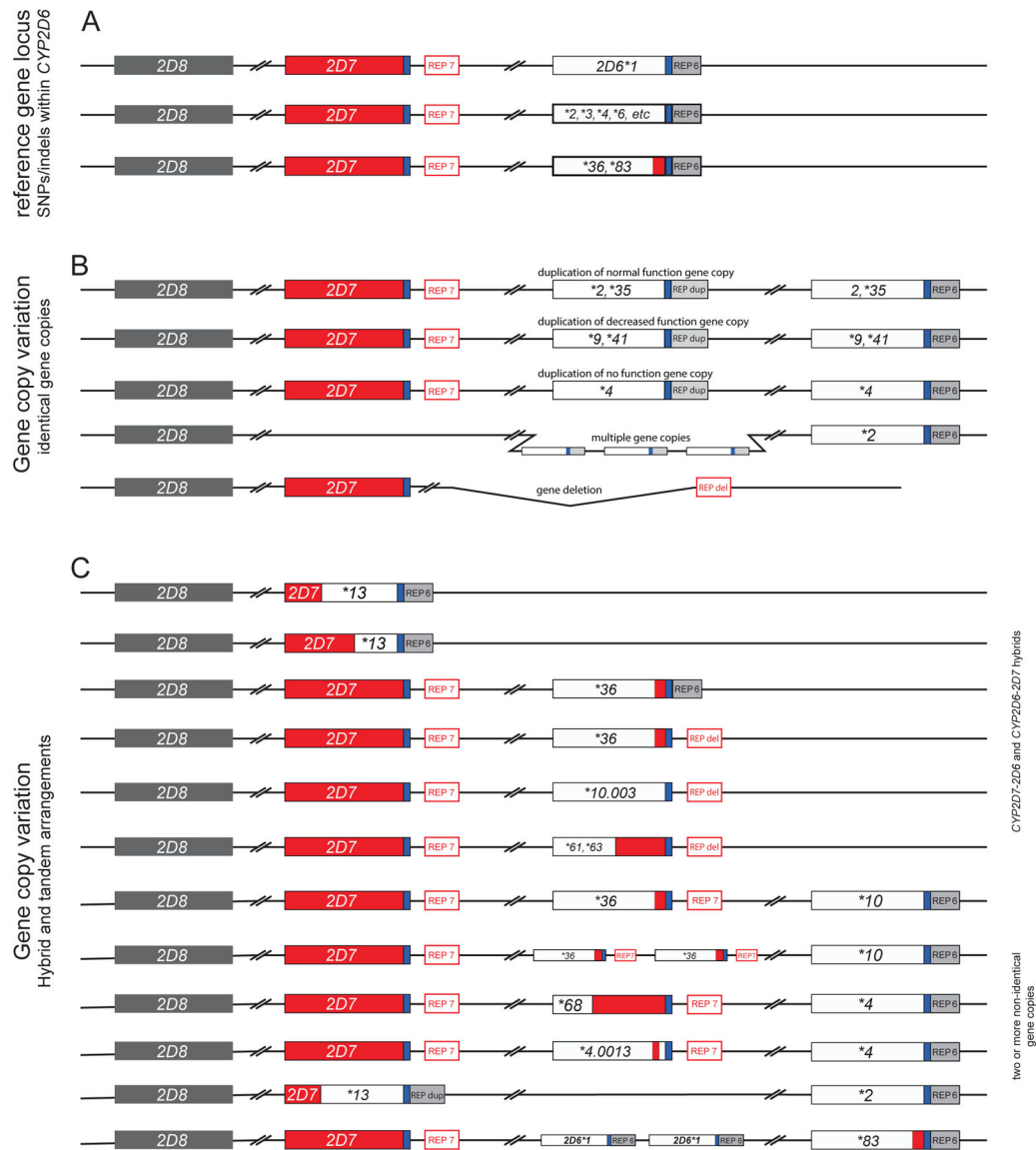
- (44). Caudle KE et al. Standardizing terms for clinical pharmacogenetic test results: consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC). *Genet Med* 19, 215–23 (2017). [PubMed: 27441996]
- (45). Gaedigk A, Simon SD, Pearce RE, Bradford LD, Kennedy MJ & Leeder JS The CYP2D6 activity score: translating genotype information into a qualitative measure of phenotype. *Clin Pharmacol Ther* 83, 234–42 (2008). [PubMed: 17971818]
- (46). Gaedigk A, Dinh JC, Jeong H, Prasad B & Leeder JS Ten Years' Experience with the CYP2D6 Activity Score: A Perspective on Future Investigations to Improve Clinical Predictions for Precision Therapeutics. *J Pers Med* 8 (2), pii: E15 (2018). [PubMed: 29673183]
- (47). PharmGKB Gene-Specific Information Tables for CYP2D6. <https://www.pharmgkb.org/page/cyp2d6RefMaterials>
- (48). Caudle KE et al. Standardizing CYP2D6 Genotype to phenotype translation: Consensus recommendations from CPIC and DPWG. CTS (2019). doi: 10.1111/cts.12692
- (49). CPIC Genotype to Phenotype Standardization Project. <https://cpicpgx.org/resources/cyp2d6-genotype-to-phenotype-standardization-project/>
- (50). Tamargo J et al. Gender differences in the effects of cardiovascular drugs. *European Heart Journal - Cardiovascular Pharmacotherapy* 3, 163–82 (2017). [PubMed: 28329228]
- (51). Ryu RJ et al. Pharmacokinetics of metoprolol during pregnancy and lactation. *J Clin Pharmacol* 56, 581–9 (2016). [PubMed: 26461463]
- (52). Blake MJ et al. Ontogeny of dextromethorphan O- and N-demethylation in the first year of life. *Clin Pharmacol Ther* 81, 510–6 (2007). [PubMed: 17301735]
- (53). Gaedigk A Complexities of CYP2D6 gene analysis and interpretation. *Int Rev Psychiatry* 25, 534–53 (2013). [PubMed: 24151800]
- (54). Shen H et al. Comparative metabolic capabilities and inhibitory profiles of CYP2D6.1, CYP2D6.10, and CYP2D6.17. *Drug Metab Dispos* 35, 1292–300 (2007). [PubMed: 17470523]
- (55). Marcucci KA, Pearce RE, Crespi C, Steimel DT, Leeder JS & Gaedigk A Characterization of cytochrome P450 2D6.1 (CYP2D6.1), CYP2D6.2, and CYP2D6.17 activities toward model CYP2D6 substrates dextromethorphan, bupropion, and debrisoquine. *Drug Metab Dispos* 30, 595–601 (2002). [PubMed: 11950793]
- (56). Bapiro T, Hasler J, Ridderström M & Masimirembwa C The molecular and enzyme kinetic basis for the diminished activity of the cytochrome P450 2D6.17 (CYP2D6.17) variant. Potential implications for CYP2D6 phenotyping studies and the clinical use of CYP2D6 substrate drugs in some African populations. *Biochem Pharmacol* 64, 1387–98 (2002). [PubMed: 12392820]
- (57). Zhou Y, Fujikura K, Mkrтчian S & Lauschke VM Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data. *Front Pharmacol* 9, 1437 (2018). [PubMed: 30564131]
- (58). Zhou Y, Mkrтчian S, Kumondai M, Hiratsuka M & Lauschke VM An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J* 19, 115–26 (2019). [PubMed: 30206299]
- (59). Hertz DL et al. In vivo assessment of the metabolic activity of CYP2D6 diplotypes and alleles. *Br J Clin Pharmacol* 80, 1122–30 (2015). [PubMed: 25907378]
- (60). Cai W-M et al. CYP2D6 genetic variation in healthy adults and psychiatric African-American subjects: implications for clinical practice and genetic testing. *Pharmacogenomics J* 6, 343–50 (2006). [PubMed: 16550211]
- (61). Zhang H, De T, Zhong Y & Perera MA The advantages and challenges of diversity in Pharmacogenomics: Can minority populations bring us closer to implementation? *Clin Pharmacol Ther* 106, 338–49 (2019). [PubMed: 31038731]
- (62). Sirugo G, Williams SM & Tishkoff SA The Missing Diversity in Human Genetic Studies. *Cell* 177, 1080 (2019). [PubMed: 31051100]
- (63). Cavallari LH et al. Multi-site investigation of strategies for the clinical implementation of CYP2D6 genotyping to guide drug prescribing. *Genet Med*, Epub 3 21 (2019). doi: 10.1038/s41436-019-0484-3

- (64). Hicks JK, Dunnenberger HM, Gumper K, Haidar CE & Hoffman JM Integrating pharmacogenomics into electronic health records with clinical decision support. *Am J Health Syst Pharm* 73, 1967–76 (2016). [PubMed: 27864204]
- (65). Caraballo PJ, Bielinski SJ, St Sauver JL & Weinshilboum RM Electronic Medical Record-Integrated Pharmacogenomics and Related Clinical Decision Support Concepts. *Clin Pharmacol Ther* 102, 254–64 (2017). [PubMed: 28390138]
- (66). Bousman CA, Jaksa P & Pantelis C Systematic evaluation of commercial pharmacogenetic testing in psychiatry: a focus on CYP2D6 and CYP2C19 allele coverage and results reporting. *Pharmacogenet Genomics* 27, 387–93 (2017). [PubMed: 28777243]
- (67). Bousman CA, Zierhut H & Muller DJ Navigating the Labyrinth of Pharmacogenetic Testing: A Guide to Test Selection. *Clin Pharmacol Ther* 106, 309–12 (2019). [PubMed: 31004441]
- (68). Pratt VM et al. Recommendations for Clinical CYP2C9 Genotyping Allele Selection: A Joint Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn* 21, 746–55 (2019). [PubMed: 31075510]
- (69). Pratt VM et al. Recommendations for Clinical CYP2C19 Genotyping Allele Selection: A Report of the Association for Molecular Pathology. *J Mol Diagn* 20, 269–76 (2018). [PubMed: 29474986]
- (70). Gaedigk A, Whirl-Carrillo M, Pratt VM, Miller NA & Klein TE PharmVar and the Landscape of Pharmacogenetic Resources *Clin Pharm Ther* accepted.
- (71). Gaedigk A, Gotschall RR, Forbes NS, Simon SD, Kearns GL & Leeder JS Optimization of cytochrome P4502D6 (CYP2D6) phenotype assignment using a genotyping algorithm based on allele frequency data. *Pharmacogenetics* 9, 669–82 (1999). [PubMed: 10634130]
- (72). Gaedigk A et al. Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events. *Clin Pharmacol Ther* 81, 242–51 (2007). [PubMed: 17259947]
- (73). Gaedigk A et al. SNP genotyping using TaqMan technology: the CYP2D6\*17 assay conundrum. *Sci Rep* 5, 9257 (2015). [PubMed: 25788121]
- (74). Gaedigk A, Garcia-Ribera C, Jeong HE, Shin JG & Hernandez-Sanchez JT Resolution of a clinical AmpliChip CYP450 Test no call: discovery and characterization of novel CYP2D6\*1 haplotypes. *Pharmacogenomics* 15, 1175–84 (2014). [PubMed: 25141893]
- (75). Gaedigk A et al. Discovery of the nonfunctional CYP2D6\*31 allele in Spanish, Puerto Rican, and US Hispanic populations. *Eur J Clin Pharmacol* 66, 859–64 (2010). [PubMed: 20473659]
- (76). Gaedigk A, Riffel AK, Berrocal BG, Solaesa VG, Davila I & Isidoro-Garcia M Characterization of a complex CYP2D6 genotype that caused an AmpliChip CYP450 Test(R) no-call in the clinical setting. *Clin Chem and Lab Med* 52, 799–807 (2014). [PubMed: 24445243]
- (77). Gaedigk A, Riffel AK & Leeder JS CYP2D6 Haplotype Determination Using Long Range Allele-Specific Amplification: Resolution of a Complex Genotype and a Discordant Genotype Involving the CYP2D6\*59 Allele. *J Mol Diagn* 17, 740–8 (2015). [PubMed: 26335396]
- (78). Rasmussen HB & Werge T Novel variant of CYP2D6\*6 is undetected by a commonly used genotyping procedure. *Pharmacol Rep* 63, 1264–6 (2011). [PubMed: 22180372]
- (79). Riffel AK et al. CYP2D7 Sequence Variation Interferes with TaqMan CYP2D6 (\*) 15 and (\*) 35 Genotyping. *Front Pharmacol* 6, 312 (2016). [PubMed: 26793106]
- (80). Black JL, Walker DL, O’Kane DJ & Harmandayan M Frequency of Undetected CYP2D6 Hybrid Genes in Clinical Samples: Impact on Phenotype Prediction. *Drug Metab Dispos* 40, 111–9 (2012). [PubMed: 22004686]
- (81). Del Tredici AL et al. Frequency of CYP2D6 Alleles Including Structural Variants in the United States. *Front Pharmacol* 9, 305 (2018). [PubMed: 29674966]
- (82). Gaedigk A, Bradford LD, Alander SW & Leeder JS CYP2D6\*36 gene arrangements within the CYP2D6 locus: association of CYP2D6\*36 with poor metabolizer status. *Drug Metab Dispos* 34, 563–9 (2006). [PubMed: 16415111]
- (83). Gaedigk A, Fuhr U, Johnson C, Berard LA, Bradford D & Leeder JS CYP2D7–2D6 hybrid tandems: identification of novel CYP2D6 duplication arrangements and implications for phenotype prediction. *Pharmacogenomics* 11, 43–53 (2010). [PubMed: 20017671]

- (84). Gaedigk A, Hernandez J, Garcia-Solaesa V, Sanchez S & Isidoro-Garcia M Detection and characterization of the CYP2D6\*9×2 gene duplication in two Spanish populations: resolution of AmpliChip CYP450 test no-calls. *Pharmacogenomics* 12, 1617–22 (2011). [PubMed: 22044417]
- (85). Gaedigk A et al. Identification of Novel CYP2D7–2D6 Hybrids: Non-Functional and Functional Variants. *Front Pharmacol* 1, 121 (2010). [PubMed: 21833166]
- (86). Gaedigk A, Twist GP & Leeder JS CYP2D6, SUL1A1 and UGT2B17 copy number variation: quantitative detection by multiplex PCR. *Pharmacogenomics* 13, 91–111 (2012). [PubMed: 22111604]
- (87). Hosono N et al. CYP2D6 genotyping for functional-gene dosage analysis by allele copy number detection. *Clinical chemistry* 55, 1546–54 (2009). [PubMed: 19541866]
- (88). Ishiguro A, Kubota T, Sasaki H & Iga T A long PCR assay to distinguish CYP2D6\*5 and a novel CYP2D6 mutant allele associated with an 11-kb EcoRI haplotype. *Clin Chim Acta* 347, 217–21 (2004). [PubMed: 15313161]
- (89). Kramer WE et al. CYP2D6: novel genomic structures and alleles. *Pharmacogenet Genomics* 19, 813–22 (2009). [PubMed: 19741566]
- (90). Huddart R et al. Standardized Biogeographic Grouping System for Annotating Populations in Pharmacogenetic Research. *Clin Pharmacol Ther* 105, 1256–62 (2019). [PubMed: 30506572]
- (91). Friedrich DC et al. Distribution of CYP2D6 Alleles and Phenotypes in the Brazilian Population. *PLoS one* 9, e110691 (2014). [PubMed: 25329392]
- (92). Dodgen TM et al. Introduction of the AmpliChip CYP450 Test to a South African cohort: a platform comparative prospective cohort study. *BMC Med Genet* 14, 20 (2013). [PubMed: 23356658]
- (93). Montane Jaime LK, Lalla A, Steimer W & Gaedigk A Characterization of the CYP2D6 gene locus and metabolic activity in Indo- and Afro-Trinidadians: discovery of novel allelic variants. *Pharmacogenomics* 14, 261–76 (2013). [PubMed: 23394389]
- (94). Zhang F & Finkelstein J Inconsistency in race and ethnic classification in pharmacogenetics studies and its potential clinical implications. *Pharmacogenomics and Personalized Medicine* 12, 107–23 (2019).
- (95). Wang D, Papp AC & Sun X Functional characterization of CYP2D6 enhancer polymorphisms. *Hum Mol Genet* 24, 1556–62 (2015). [PubMed: 25381333]
- (96). PharmVar Standards document.
- (97). PharmVar Allele Designation and Evidence Level document. [www.pharmvar.org/documents/Allele\\_Designation\\_and\\_Evidence\\_Level\\_Criteria\\_V2.1.pdf](http://www.pharmvar.org/documents/Allele_Designation_and_Evidence_Level_Criteria_V2.1.pdf)
- (98). PharmVar CYP2D6 Gene Expert Panel roster. <https://www.pharmvar.org/expert-panels>
- (99). PharmVar P450 Nomenclature site – Archive [https://www.pharmvar.org/htdocs/archive/index\\_original.htm](https://www.pharmvar.org/htdocs/archive/index_original.htm); previously at [cypalleles.ki.se](http://cypalleles.ki.se)
- (100). NCBI Reference Sequences database. <https://www.ncbi.nlm.nih.gov/refseq/>
- (101). Locus Reference Genomic (LRG) project. <https://www.lrg-sequence.org/>
- (102). PharmVar CYP2D6 gene page. <https://www.pharmvar.org/gene/CYP2D6>
- (103). Gaedigk A et al. The Evolution of PharmVar. *Clin Pharmacol Ther* 105, 29–32 (2019). [PubMed: 30536702]
- (104). Pratt VM et al. Characterization of 137 Genomic DNA Reference Materials for 28 Pharmacogenetic Genes: A GeT-RM Collaborative Project. *J Mol Diagn* 18, 109–23 (2016). [PubMed: 26621101]
- (105). Gaedigk A et al. Characterization of Reference Materials for Genetic Testing of CYP2D6 Alleles: A GeT-RM Collaborative Project. *J Mol Diagn*, Epub 8 9 (2019). doi:10.1016/j.jmoldx.2019.06.007
- (106). Twist GP et al. Constellation: a tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene, CYP2D6, from whole-genome sequences. *NPJ Genom Med* 1, 15007 (2016). [PubMed: 29263805]
- (107). Lee SB et al. Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. *Genet Med* 21, 361–72 (2019). [PubMed: 29875422]

- (108). Qiao W, Wang J, Pullman BS, Chen R, Yang Y & Scott SA The CYP2D6 VCF Translator. *Pharmacogenomics J* 17, 301–3 (2017). [PubMed: 26975230]
- (109). Numanagic I et al. Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. *Nat Commun* 9, 828 (2018). [PubMed: 29483503]
- (110). Numanagic I, Malikić S, Pratt VM, Skaar TC, Flockhart DA & Sahinalp SC Cypiripi: exact genotyping of CYP2D6 using high-throughput sequencing data. *Bioinformatics* 31, i27–34 (2015). [PubMed: 26072492]
- (111). Cohn I et al. Genome sequencing as a platform for pharmacogenetic genotyping: a pediatric cohort study. *npj Genomic Medicine* 2, 19 (2017). [PubMed: 29263831]
- (112). 1000 Genomes Project; uncallable regions. <http://www.internationalgenome.org/data/>; <https://www.nature.com/articles/nature15393> available in Suppl materials section 9.2
- (113). Scantamburlo G et al. Allele Drop Out Conferred by a Frequent CYP2D6 Genetic Variation For Commonly Used CYP2D6\*3 Genotyping Assays. *Cell Physiol Biochem* 43, 2297–309 (2017). [PubMed: 29073588]
- (114). Qiao W et al. Long-Read Single Molecule Real-Time Full Gene Sequencing of Cytochrome P450–2D6. *Hum Mutat* 37, 315–23 (2016). [PubMed: 26602992]
- (115). Buermans HP et al. Flexible and Scalable Full-Length CYP2D6 Long Amplicon PacBio Sequencing. *Hum Mutat* 38, 310–6 (2017). [PubMed: 28044414]
- (116). Liao Y, Maggo S, Miller AL, Pearson JF, Kennedy MA & Cree SL Nanopore sequencing of the pharmacogene CYP2D6 allows simultaneous haplotyping and detection of duplications. *Pharmacogenomics* accepted, (2019).
- (117). 10X Genomics (Linked Reads Genomics). <https://www.10xgenomics.com/>





**Figure 1. Selection of structural variants illustrating the complexity of *CYP2D6* gene copy number variation.**

The top line in Panel A depicts the reference gene locus containing a single copy of the *CYP2D6* gene. The 2<sup>nd</sup> line represents *CYP2D6*\*2, \*3, \*4, \*6, etc. Panel B exemplifies allelic variants carrying two or more (multiple) normal function (e.g. \*2xN, \*35xN), decreased function (e.g. \*9xN, \*41xN) or nonfunctional (e.g. \*4xN) gene copies. The duplicated and multiplied gene copies shown in this panel are believed to be identical. The last line in B represents the *CYP2D6*\*5 gene deletion allele. Of note, alleles with duplicated gene copies have a *CYP2D6*-like REP-dup sequence without the 1.6 kb long *CYP2D7* spacer sequence. Panel C depicts the most complex structural variants. These harbor a singleton hybrid gene or carry two or more non-identical gene copies. The duplicated gene in such arrangements often, but not always, has a *CYP2D7*-like downstream region including the 1.6 kb long spacer sequence. *CYP2D6* and *CYP2D7*-derived sequences are shown in white/gray and red, respectively. A common downstream element is shown in blue

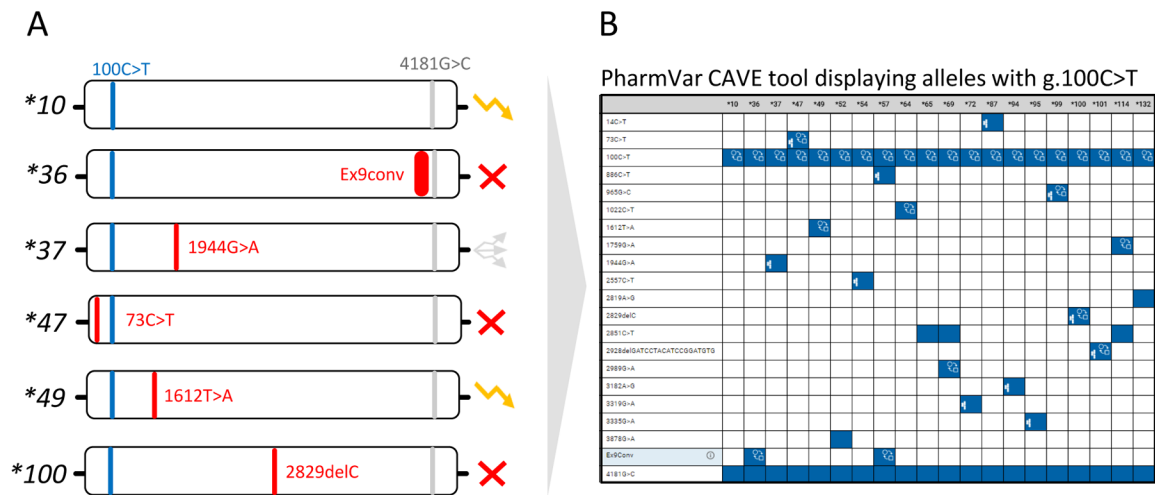
and repetitive elements (REP6, REP7, REPdup, and REPdel) are gray or red based on whether they resemble *CYP2D6* (without the spacer) or *CYP2D7* (with the spacer).

Author Manuscript

Author Manuscript

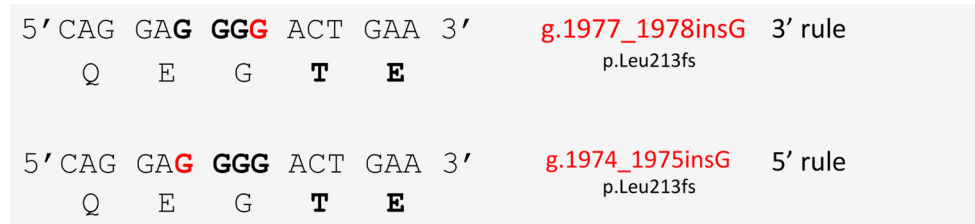
Author Manuscript

Author Manuscript



### Figure 2. Allele default assignment strategy used by many testing platforms

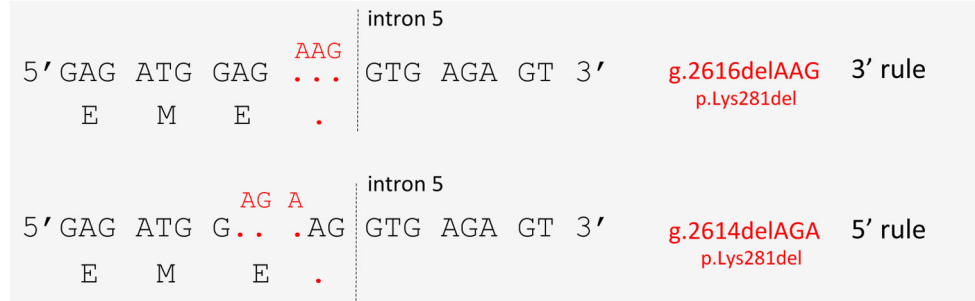
Many platforms test for a panel of the more commonly observed SNVs, but not all known SNVs or all alleles defined by PharmVar. As a consequence, allele assignments are made by ‘default’ as exemplified on those that are defaulted to *CYP2D6*\*10. It is imperative to know which SNVs are tested in order to garner a full understanding of how phenotype is derived as well as to fully understand a test’s limitations. Panel **A** depicts a selection of allelic variants all carrying g.100C>T. An unequivocal *CYP2D6*\*10 call can only be made after ruling out the presence of numerous other SNVs, e.g. those defining \*36, \*37, \*47, \*49, \*100 and others. These alleles can be nonfunctional, have decreased or uncertain function as indicated by the function symbols, or await function assignment by CPIC in which case alleles are labeled as ‘awaits curation’ on the PharmVar *CYP2D6* gene page. Panel **B** shows all currently defined alleles containing the g.100C>T core SNV (see Figure 4 and text for more information regarding core allele definitions). The depicted graph was generated with the PharmVar’s CAVE tool.

**A** *CYP2D6*\*20 (rs72549354)

\*20 (rs72549354)

	PharmVar annotation	dbSNP annotation
SNV type	indel	
Sequence Start	g.6177_6178insG	g.6177dup
ATG start	g.1977_1978insG	n/a

rs72549354 is shown on dbSNP as CCC&gt;CCCC (or GGG&gt;GGGG)

**B** *CYP2D6*\*9 (rs5030656)

\*9 (rs5030656)

	PharmVar annotation	dbSNP annotation
SNV type	indel	
Sequence Start	g.6816delAAG	g.6816_6818delAAG
ATG start	g.2616delAAG	n/a

rs5030656 is shown on dbSNP as CTTCT&gt;CT (or AGAAG&gt;AG)

**Figure 3. SNV alignment and Coordinates**

Panel **A** shows *CYP2D6*\*20 which is characterized by a g.1977\_1978insG (rs72549354) (red), which causes a frameshift at amino acid position 213 that obliterates function. This SNV is embedded within three 'G' (bold) that are flanked on each side by one 'A' (5' AGGGA 3'). Because of the nature of the surrounding sequence, the actual insertion site is unclear. PharmVar displays the position of this variant as g.1977\_1978 (per the 3' rule) as opposed to g.1974\_1975 (5' rule). Of note, dbSNP uses the same position (g.6177 counting from sequence start) for this variant for the RefSeq annotation, but describes the 'insertion' of the 'G' as 'duplication'.

Panel **B** shows *CYP2D6*\*9 which is characterized by a 3-base pair deletion (red) commonly described as g.2616delAAG (rs5030656) that results in the loss of a lysine (K281del) causing decreased function. Because of the nature of the surrounding sequence it remains unclear, however, which three bases are deleted. PharmVar displays the position of this in-

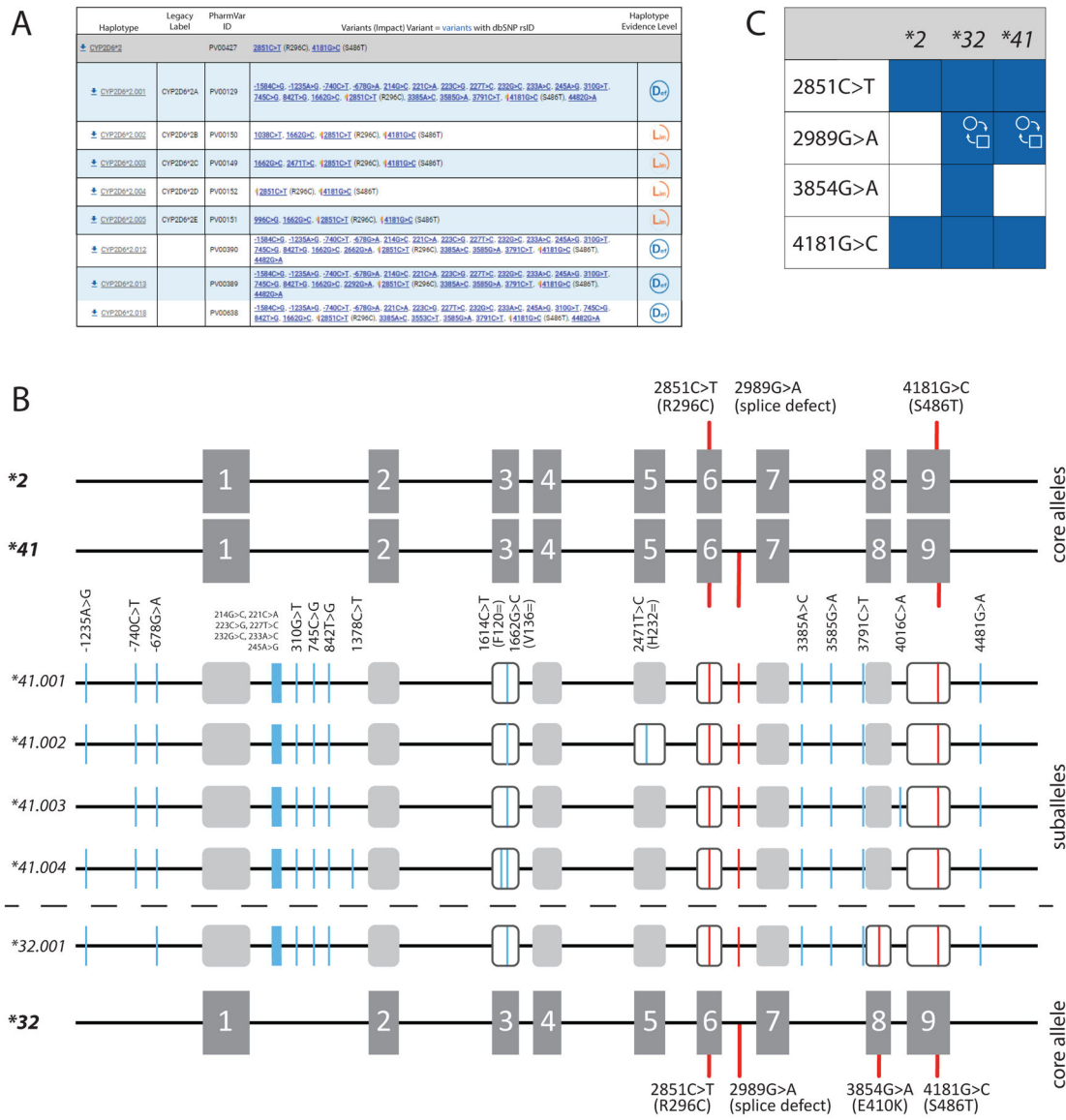
frame deletion as g.2616delAAG (per the 3' rule) as opposed to g.2614delAGA (per the 5' rule). Of note, dbSNP uses the same position (g.6816 counting from the sequence start) for this variant for the RefSeq annotation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 4. Overview of core alleles, suballeles and the graphical Core Allele ViewEr (CAVE)**  
 Panel A shows the *CYP2D6*\*2 core allele definition (gray bar). Core SNVs and the allele's PharmVar ID (PVID) are as shown. A selection of suballeles is provided under the core allele definition (as of July 2019, there were 20 *CYP2D6*\*2 suballeles). Legacy allele designations are cross-referenced (e.g. \*2.001 corresponds to \*2A). Two of the suballeles depicted, \*2.002 and \*2.003 have been defined by exon sequencing only and therefore are assigned a 'Lim' evidence level. Panel B is a graphical representation of the *CYP2D6*\*2, \*32 and \*41 core alleles and their core SNVs. g.2851C>T (R296C) and g.4181G>C (S486T) are present on all three, while g.2989G>A (splice defect) is present on *CYP2D6*\*32 and \*41, and g.3854G>A (E410K) is only found on \*32. Gray boxes represent the nine exons. The middle section shows four of the five *CYP2D6*\*41 suballeles defined to date. Core SNVs (causing an amino acid change or aberrant splicing) are shown in red, all other SNVs are highlighted in blue; core SNVs are found on all suballeles. Panel C depicts the CAVE output

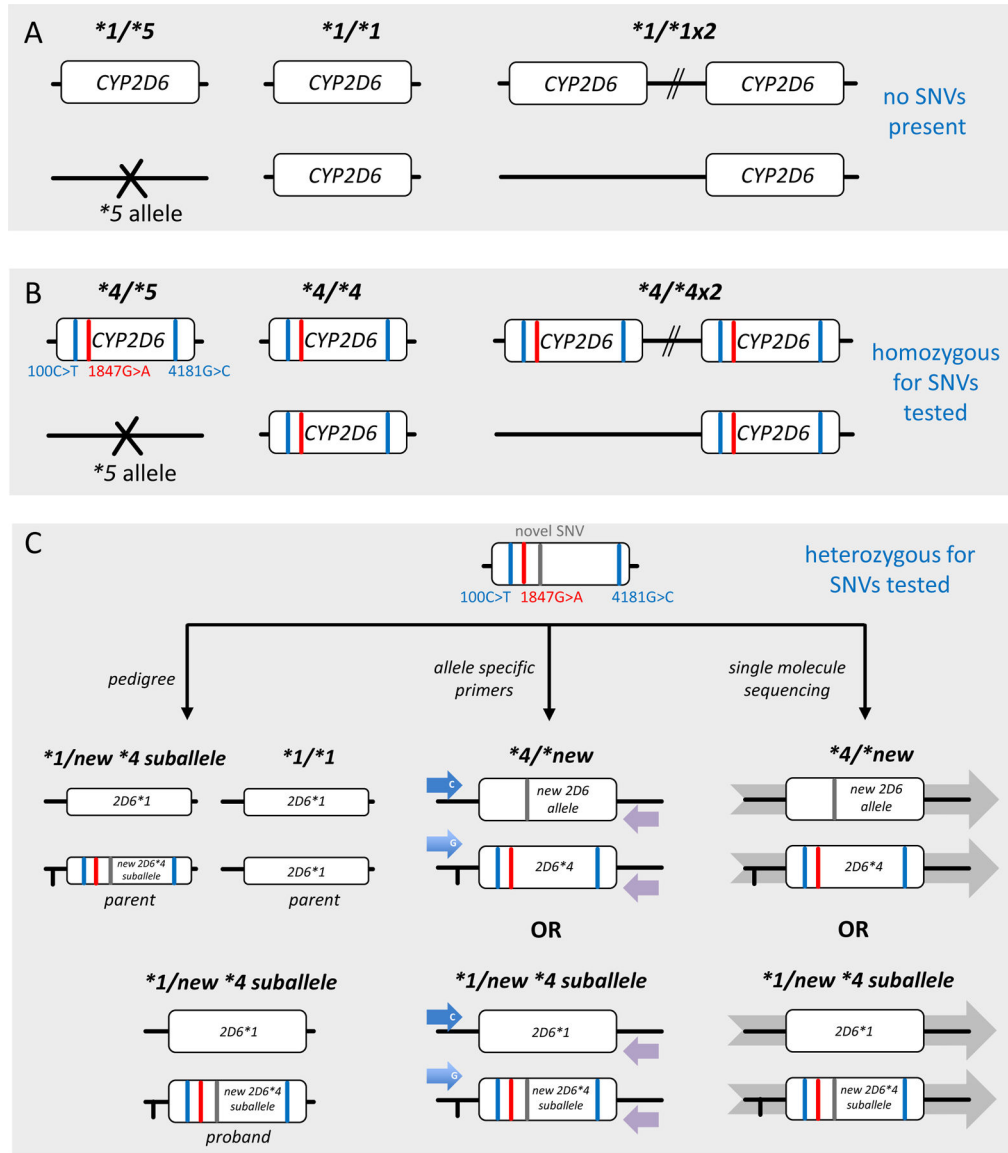
visualizing the core SNVs shared among *CYP2D6*\*2, \*32 and \*41. Blue boxes indicate the presence of a SNV and the function symbol (Ⓜ) indicates that g.2989G>A alters function. It remains unknown, however, whether 3854G>A (E410K) exerts a functional impact.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 5. Experimental approaches for phasing SNVs**

Panel A shows the *CYP2D6* reference gene locus, i.e. no SNVs are present. This subject has a *CYP2D6*\*1/\*1 genotype in the absence of CNVs. If CNV testing yields 1 and 3 copies, genotypes will be assigned as *CYP2D6*\*1/\*5 and \*1×2/\*1, respectively. Panel B shows an example which is homozygous for three SNVs (depicted as red or blue lines), i.e. each SNV is present on each gene copy. This subject has a *CYP2D6*\*4/\*4 genotype in the absence of CNVs. If CNV testing yields 1 and 3 copies, genotypes will be assigned as *CYP2D6*\*4/\*5 and \*4×2/\*4 (or alternatively \*4×3/\*5, not shown). Panel C shows a sample that is heterozygous for three SNVs (same as in B) and a novel SNV. It is impossible, however, to know whether the novel SNV is in *cis* (on the same allele as other three SNVs signifying *CYP2D6*\*4) or in *trans* (by itself on the opposite allele). The bottom panel visualizes different approaches of how haplotype can be inferred, e.g. inheritance (left-hand graph) or



experimentally determined, e.g. allele-specific long-range PCR followed by sequencing (center graph) or single molecule sequencing (right-hand graph).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**

Links to Sites and Online Resources Referenced Throughout the Review

Sources and References	References
<b>PharmVar</b>	
<i>CYP2D6</i> important gene information <ul style="list-style-type: none"> <li>Gene Read Me document</li> <li>Change Log document</li> <li>Structural Variation document</li> </ul>	(102)
Standards document	(96)
Allele Designation and Evidence Level document	(97)
<i>CYP2D6</i> Gene Expert Panel roster	(98)
P450 Nomenclature site – Archive	(99)
<b>PharmGKB</b>	
<i>CYP2D6</i> gene page	(43)
Gene-Specific Information Tables for <i>CYP2D6</i> <ul style="list-style-type: none"> <li>Allele Definition Table</li> <li>Allele Functionality Table</li> <li>Frequency Table</li> <li>Diplotype-Phenotype Table</li> <li>Gene Resource Mappings</li> </ul>	(47)
<ul style="list-style-type: none"> <li><i>CYP2D6</i> Drug Label Annotations</li> </ul>	(22)
<b>CPIC</b>	
Guidelines	(42)
Gene/drug pairs <ul style="list-style-type: none"> <li>process for assigning CPIC levels</li> <li>levels for gene/drug pairs</li> <li>process for prioritizing CPIC guidelines</li> </ul>	(24)
Genotype to Phenotype Standardization Project	(49)
<b>Other Resources</b>	
Drug Interactions Flockhart Table™	(31)
FDA Pharmacogenomic Biomarkers in Drug Labeling	(23)
NCBI Reference Sequences database	(100)
Locus Reference Genomic (LRG) project	(101)
1000 Genomes Project and callable genome mask materials	(112)
10X Genomics (linked Reads Genomics)	(117)

Table 2

Novel allele(s) and suballele(s) now fully defined

Core Allele Designation	Novel alleles	number of new alleles
*115 - *139	*115.001 - *139.001	25
	Alleles now defined by full-length sequence definition <sup>1</sup>	
	*1.002, *1.005, *2.001, *2.005, *3.001, *4.001, *4.004, *6.001, *6.002, *6.004, *7.001, *9.001, *11.001, *14.001, 15.001, *17.001, *22.001, *28.001, *29.001, *32.001, *33.001, *35.001, *39.001, *40.001, *41.001, *42.001, *43.001, *58.001, *59.001, *90.001, *106.001, *112.001 and *113.001	
	Novel suballeles	
*1	*1.006 - *1.032	27
*2	*2.012 - *2.020	9
*4	*4.015 - *4.028	14
*6	*6.005, *6.006	2
*9	*9.002	1
*12	*12.002	1
*15	*15.002, *15.003	2
*17	*17.002, *17.003	2
*28	*28.002	1
*35	*35.003 - *35.007	5
*36	*36.002	1
*41	*41.002 - *41.005	4
*43	*43.002	1
*46	*46.003	1
*52	*52.002	1
*56	*56.003	1
*71	*71.002, *71.003	2
*83	*83.002, *83.003	2
*84	*84.002	1
*106	*106.002	1
*111	*111.002	1

<sup>1</sup>Evidence levels for these alleles were revised from 'Lim' (Limited) or 'Mod' (Moderate) to 'Def' (Definitive)

As of September 2019 there are 131 *CYP2D6* core allele definitions (\*1 through \*139, eight alleles have been retired).

**Table 3**

Summary of edits and changes made as alleles were transitioned into the PharmVar database

Reason	Change	Affected Alleles
Standardization	<i>CYP2D7</i> intron 1 and exon 9 conversions are now specified	<i>*2.001, *2.011, *4.013, *11.002, *14.001, *21.002, *31.001, *35.002, *36.001, *41.001, *51.001, *56.001, *57.001, *58.001, *73.001, *83.001, *84.001, *88.001, *98.001, *102.001-105.001</i>
	Comment 'variable number of A's in the region -1258 to -1237' removed	<i>*2.011, *10.002, *21.001, *36.001, *45.001</i>
	Variants beyond mapped gene regions removed	<i>*11.002, *31.001, *35.002, *73.001, *84.001, *85.001</i>
	Position edits or corrections	<i>*2.011, *18.001, *29.001, *40.001, *46.001, *47.001, *58.001, *83.001, *84.001, *111.001</i>
	Positions changed due to 3'-Alignment rule	<i>*20.001, *21.001, *21.002, *42.001</i>
Other	Retired star alleles removed	<i>*2J, *10D</i>
	Comments removed	<i>*4.013, *57.001, *82.001</i>