



Published in final edited form as:

*Clin Pharmacol Ther.* 2021 February ; 109(2): 352–366. doi:10.1002/cpt.1973.

## PharmVar GeneFocus: *CYP2C19*

Mariana R. Botton<sup>1</sup>, Michelle Whirl-Carrillo<sup>2</sup>, Andria L Del Tredici<sup>3</sup>, Katrin Sangkuhl<sup>2</sup>, Larisa H. Cavallari<sup>4</sup>, José A. G. Agúndez<sup>5</sup>, Jorge Duconge<sup>6</sup>, Ming Ta Michael Lee<sup>7</sup>, Erica L. Woodahl<sup>8</sup>, Karla Claudio-Campos<sup>4</sup>, Ann K. Daly<sup>9</sup>, Teri E. Klein<sup>2</sup>, Victoria M. Pratt<sup>10</sup>, Stuart A. Scott<sup>11,\*</sup>, Andrea Gaedigk<sup>12,\*</sup>

<sup>1</sup>Hospital de Clínicas de Porto Alegre, Brazil

<sup>2</sup>Department of Biomedical Data Science, Stanford University, Stanford, CA, USA

<sup>3</sup>Millennium Health LLC, San Diego, CA, USA

<sup>4</sup>University of Florida, Gainesville, FL, USA

<sup>5</sup>University Institute of Molecular Pathology Biomarkers, UNEx, ARADyAL, Instituto de Salud Carlos III, Cáceres, Spain

<sup>6</sup>University of Puerto Rico, School of Pharmacy, Medical Sciences Campus, San Juan, PR

<sup>7</sup>Genomic Medicine Institute, Geisinger, Danville, USA

<sup>8</sup>Department of Biomedical and Pharmaceutical Sciences, University of Montana, Missoula, MT, USA

<sup>9</sup>Newcastle University, Newcastle upon Tyne, England, UK

<sup>10</sup>Indiana University School of Medicine, Indianapolis IN, USA

<sup>11</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA and Sema4, Stamford, CT, USA

<sup>12</sup>Division of Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy, Kansas City, MO, USA

### Abstract

The Pharmacogene Variation Consortium (PharmVar) catalogues star (\*) allele nomenclature for the polymorphic human *CYP2C19* gene. *CYP2C19* genetic variation impacts the metabolism of many drugs and has been associated with both, efficacy and safety issues for several commonly prescribed medications. This GeneFocus provides a comprehensive overview and summary of *CYP2C19* and describes how haplotype information catalogued by PharmVar is utilized by the

---

marirb@gmail.com.

\*co-senior authors

Author contributions

All authors wrote the manuscript.

Conflicts of Interest:

Indiana University School of Medicine Pharmacogenomics Laboratory, Millennium Health, and Sema4 are fee-for-service clinical laboratories that offer clinical pharmacogenetic testing. A.L.D. is a paid employee of Millennium Health; S.A.S., is a paid employee of Sema4. All other authors declared no competing interests for this work.

## CYP2C19 Brief History

A novel drug hydroxylation polymorphism in *CYP2C19* was first described in 1984 after deficient metabolism of the anticonvulsant drug mephenytoin was identified in healthy volunteers and in family studies (1). Mephenytoin was subsequently used as a probe drug to distinguish between normal and poor metabolizers, which ultimately was determined to be due to variable expression of cytochrome P450 2C19 (*CYP2C19*) (2, 3). The *CYP2C19* gene was first characterized in 1991 (4), and is one of four *CYP2C* subfamily genes, located on chromosome 10q23.33 (*CYP2C8*, *CYP2C9*, *CYP2C18* and *CYP2C19*) (5). The first *CYP2C19* variant allele, implicated in poor metabolism (*CYP2C19\*2*), was reported in 1994 (6). Numerous variants were subsequently reported and continue to be identified (7, 8). While many *CYP2C19* variants have been reported as having no function, a notable discovery in *CYP2C19* pharmacogenetics was the identification of the *CYP2C19\*17* allele, which is associated with increased enzyme activity and more rapid drug clearance (9). A systematic nomenclature system for *CYP2C19* alleles was first proposed in 2000 (10) and continues to be updated as new alleles are described. Another two milestones in *CYP2C19* history are the FDA revision of drug labeling (e.g., clopidogrel with boxed warning added) based on *CYP2C19* pharmacogenetics and the evidence-based CPIC guidelines issued for various *CYP2C19* gene-drug pairs.

## Status of Nomenclature before PharmVar

Star (\*) nomenclature was first applied to *CYP2D6* in the 1990s (11), which then expanded to include *CYP2C19* and other cytochrome P450 (*CYP*) genes (10). *CYP* star (\*) allele nomenclature was maintained by the Human Cytochrome P450 Allele Nomenclature Database (12), which was widely accepted by the pharmacogenetics community as the central resource for cataloguing *CYP* variation. Thirty-five *CYP2C19* haplotypes, *CYP2C19\*1* through *\*35*, were submitted by investigators to this database before it was transitioned to the Pharmacogene Variation (PharmVar) Consortium in 2017 (13). Although only exons were required to be sequenced for submission to the original *CYP* nomenclature site, upstream gene regions were eventually also interrogated, submitted, and included in haplotype definitions as investigators searched for variants in regulatory elements. Several submissions also contained intronic variants located near exon-intron boundaries, which were included in haplotype definitions, regardless of whether they were experimentally shown to impact *CYP2C19* function. So-called 'suballeles', e.g., *CYP2C19\*2A-J*, were initially also catalogued, but eventually no longer considered for independent naming.

The information provided by the *CYP* nomenclature site was utilized by knowledge resources (e.g., PharmGKB) as well as the pharmacogenetics testing and implementation communities, including clinical genetic testing laboratories and the Clinical Pharmacogenetics Implementation Consortium (CPIC) (14). As the genomics field progressed with remarkable advances, more human genetic variants were discovered,

including pharmacogenetic variants, which ultimately resulted in challenges with reconciling *CYP* variants into the star (\*) allele system. This also identified a need for a more standardized allele submission system and rigorous nomenclature protocol, as well as a more open-access haplotype definition database.

In this review, Human Genome Variation Society (HGVS) guidelines (Table 1) are used to report star (\*) allele sequence variants according to their relative position in the *CYP2C19* NM\_000769.1 transcript, with the ATG start codon being +1; corresponding protein coordinates are also provided. For example, the *CYP2C19*\*3 allele, defining variant (rs4986893), is referred to as c.636G>A (p.W212X), indicating the variant affects the nucleotide at cDNA position 636 and causes a tryptophan to stop codon change at amino acid position 212.

## Clinical Significance

The *CYP2C19* enzyme is part of the CYP450 superfamily contributing to the metabolism of many clinically used drugs, including clopidogrel, voriconazole, proton pump inhibitors, antidepressants, carisoprodol, diazepam. Moreover, *CYP2C19* also metabolizes endogenous substances such as melatonin and progesterone (8). More information on drugs metabolized by *CYP2C19* can be found in the PharmGKB drug label annotations (15), the US Food and Drug Administration (FDA) table on Pharmacogenomic Biomarkers in Drug Labeling (16), the FDA Table of Pharmacogenetic Associations (17), and the CPIC drug-gene pairs (18). Links to resources mentioned here and throughout this GeneFocus are summarized in Table 1.

One of the most common clinical applications of *CYP2C19* genotyping is to predict response to clopidogrel and guide selection of antiplatelet therapy after percutaneous coronary intervention (PCI) among patients with acute coronary syndromes (ACS) (19). Data on outcomes with this genotype-guided approach to antiplatelet therapy are beginning to emerge (20, 21). Moreover, studies in multi-ethnic, more diverse populations – including some in minority, underrepresented groups – are also underway (22). It is expected that results from these studies will contribute significantly to closing the current knowledge gap in the field regarding the identification of novel ethno-specific variants of clinical value, as well as validating previous findings from early studies in diverse populations at large.

Other clinical uses of *CYP2C19* genetic testing are to guide prescribing of selective serotonin reuptake inhibitors in the treatment of depression, dosing of proton pump inhibitors for gastroesophageal reflux disease, and dosing of voriconazole for prophylaxis of invasive fungal infection among others. A recent study of *CYP2C19*-guided voriconazole dosing for antifungal prophylaxis in adult allogeneic hematopoietic cell transplant recipients demonstrated a decrease in the percent of patients with subtherapeutic concentrations and improved treatment success rates with the genotype-guided approach compared to historical controls, with an estimated cost savings of \$4,700 per patient (23).

Given the clinical significance of *CYP2C19* genotype information, it is imperative that test information is accurate and that providers understand the limitations of a genotype test.

Recommendations for clinical *CYP2C19* genotyping allele selection have been published (24) and test options are discussed in detail (25, 26) elsewhere.

### Other Factors that can Influence CYP2C19 activity

A patient's metabolic profile may be profoundly impacted by one or multiple co-medications. The magnitude of drug-drug interaction depends on the individual metabolizer phenotype and the relative contribution of *CYP2C19* to the metabolism of a drug(s). For example, fluconazole, fluoxetine, and fluvoxamine are known inhibitors of *CYP2C19*, and may phenoconvert individuals with a normal *CYP2C19* genotype to intermediate or poor metabolizers, while rifampin, efavirenz, and St. John's wort are known inducers (27, 28). Of note, drug interactions are of no consequence for *CYP2C19* poor metabolizers who lack enzyme activity. Drug-drug interactions are of particular concern in the elderly as polypharmacy is common in this population (29). For more information on *CYP2C19* inhibitors and inducers we refer to the Drug Interactions Flockhart Table™ (30).

It has also been suggested that inflammation may influence efficacy. As shown in recent studies, chronic inflammation was associated with significantly reduced *CYP2C19* activity by 46% in type-2 diabetes patients (31). In the OPTIMUS study, activity was reported to be reduced by 60% and type-2 diabetes patients remained poor responders although maintenance doses of clopidogrel were doubled (32).

### Sex and age-related differences

Although gender-specific *CYP2C19* gene expression has been observed in mice (higher in males than females in liver and kidney), this remains controversial in human studies (33). A number of investigations found no differences between gender (34, 35), however, these were contradicted by additional investigations that reported higher *CYP2C19* activity in females (36) than males (37).

Age-related differences in *CYP2C19* gene and protein expression, as well as metabolic activity, have been reported. *CYP2C19* gene and protein expression were found to be the lowest in fetal and the highest in pediatric samples (38, 39). *CYP2C19* protein expression appears to linearly increase during the first 5 months of life but is variable from this age to 10 year of age (21-fold) (40). It was also shown that *CYP2C19* mRNA levels were lower in adult compared to pediatric samples suggesting that *CYP2C19* expression may peak during childhood (41).

### *CYP2C19* and the PharmGKB

The PharmGKB collects, curates and disseminates knowledge about the impact of human genetic variation on drug response (42). The *CYP2C19* gene page allows structured access to gene-specific pharmacogenomic 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), the Canadian

Pharmacogenomic Network for Drug Safety (CPNDS), and 2) “Rx study annotations” that provide genotype-based drug dosing, reported in individual journal articles. Nine CPIC, 19 DPWG clinical guideline annotations (10 with recommendations and 9 providing no drug-specific recommendations) and 12 “Rx study annotations” are available as of early 2020 for *CYP2C19* with overlapping *CYP2C19*-drug pairs.

PharmGKB extracts pharmacogenomic-relevant information from agency-approved drug labels and applies a “pharmacogenetics (PGx) level” category, based on the level of action implied in each label (e.g., Testing required, Actionable PGx, Testing recommended or Informative PGx). On the *CYP2C19* page, annotations can be accessed (as of early 2020) for 23 US FDA approved labels, 9 European Medicines Agency (EMA) approved labels, 6 Pharmaceuticals and Medical Devices Agency, Japan (PMDA) approved labels, 9 Health Canada (HCSC) approved labels, and 12 Swiss Agency of Therapeutic Products (Swissmedic) approved labels. Currently, PharmGKB contains 82 *CYP2C19* related clinical annotations, which are evidence-rated genotype/diplotype level summarizes for specific variant/allele-drug combinations based on curated single literature entries (variant annotations). Pharmacokinetic pathways depicting *CYP2C19* in drug metabolism are available for 30 drugs, although the significance for *CYP2C19* involvement varies by drug. PharmGKB and CPIC work collaboratively to develop gene-specific resources that accompany each CPIC guideline, including allele definition mapping, allele functionality, allele frequency, and diplotype to phenotype mapping files in a standardized format. Gene-specific information tables for *CYP2C19* are available from PharmGKB (44). In addition, the Pharmacogenomics Clinical Annotation Tool (PharmCAT) is currently under development by PharmGKB to facilitate the interpretation and reporting of pharmacogenomic-based dosing recommendations (45).

### ***CYP2C19* and the Clinical Pharmacogenetics Implementation Consortium (CPIC)**

The CPIC develops structured, evidence-based, clinical practice guidelines for drugs affected by pharmacogenetic variation (46, 47). Several *CYP2C19* gene-drug pairs have been prioritized through consideration of multiple factors such as the available body of literature, severity of clinical consequences, availability of alternative therapies and whether a prescribing change (drug choice or dose) is warranted. To date, four CPIC guidelines have been published that include *CYP2C19*: tricyclic antidepressants (48), selective serotonin reuptake inhibitors (49), clopidogrel (50) and voriconazole (51), and a *CYP2C19* guideline on proton pump inhibitors (e.g., esomeprazole, omeprazole) is currently in preparation. Each guideline has multiple components, including *CYP2C19* phenotype-specific therapeutic recommendations, systematic evidence review, and implementation resources to support the translation of the guideline into electronic health records (EHRs) with example clinical decision support (CDS) text.

To facilitate clinical implementation of pharmacogenetic information and portability of pharmacogenetic data across institutions, CPIC established standard terminology for allele functional status and phenotypes for pharmacogenes, including *CYP2C19*. CPIC specifically recommends the terms “no function,” “decreased function,” “normal function,”

and "increased function" to describe alleles, and "poor, intermediate, normal, rapid, and ultrarapid" to describe inferred drug metabolizing enzyme phenotypes (52).

## Genotype to Phenotype Translation

An individual has two *CYP2C19* haplotypes, one on each chromosome, which constitutes his/her diplotype. For example, a *CYP2C19*\*2/\*3 diplotype assignment indicates that one chromosome (or allele) has single nucleotide variations (SNVs) defining the *CYP2C19*\*2 haplotype and the second chromosome (or allele) has SNVs defining the *CYP2C19*\*3 haplotype. The term "genotype" can refer either to the sum of all detected SNVs or to a person's diplotype.

For functional classification, individuals are categorized into the five CPIC-recommended phenotype categories: poor (PM), intermediate (IM), normal (NM) (formerly EM, extensive), rapid (RM), and ultrarapid (UM) metabolizers (52). Unlike the *CYP2D6* gene, which utilizes the Activity Score system to facilitate genotype to phenotype translation (53), the *CYP2C19* gene translation process is based on the combination of increased, normal, decreased and no function alleles as provided in the Diplotype-Phenotype-Table by the PharmGKB and CPIC (44). The table provides all possible *CYP2C19* diplotypes and their respective phenotypes. An individual's predicted phenotype is based on the function of the two alleles present in an individual.

## Need for standardized genetic variation definitions and reporting of functional/clinical impact

In order to guide drug therapy, it is imperative to understand *CYP2C19* allelic variation as well as allele and genotype function. Although many alleles have been observed in phenotypic PMs and their causative genetic variations described (e.g., *CYP2C19*\*2, \*3, and \*4, etc.), the function of an allele may be unknown (currently none for *CYP2C19*) or uncertain. *In vitro* characterization of allelic variants (39, 54-58) often produce results that are inconsistent among test systems which may be attributed to differences in the substrates used and between the experimental variables (see *CYP2C19* functionality table for a detailed summary (44)). Although *in silico* prediction tools are improving, *in vivo* validation is still the gold standard. Therefore, for any given allele (except for those shown to completely abolish activity) caution should be taken, when extrapolating functional data from one drug or substrate to another. Ideally, one would be able to assess the *in vivo* function of each individual *CYP2C19* haplotype with each individual *CYP2C19* substrate, which would refine the phenotype predicting capacity of *CYP2C19* genetic testing. In addition, co-medications (drug-drug interactions) may not affect all *CYP2C19* variants equally, and there is still limited or no information regarding genetic variability for many minority populations (59, 60).

For many drug metabolizing enzymes, the combination of sequence variations that define haplotypes is critical to precisely predict enzyme function. A notable example for *CYP2C19* is c.-806C>T, which is the defining core SNV of the *CYP2C19*\*17 increased function allele but can also be found with c.1A>G on the no function *CYP2C19*\*4.002 suballele. Given that

both genotyping and short-read sequencing are unable to determine the phase of variants like c.-806C>T and c.1A>G, precisely inferring *CYP2C19* haplotypes remains a challenge.

As previously reported in the *CYP2D6* GeneFocus (61), clinical pharmacogenetic programs have successfully been implemented over the past years, but numerous challenges remain to accelerate adoption (62). Standardization is a key area that continues to represent an opportunity for all pharmacogenetic stakeholders to improve upon, including laboratory processes, test ordering, result reporting, and data representation. This is in alignment with recent reports, which emphasized that clinically actionable pharmacogenetic information must be accurately represented in electronic health records by using a harmonized system for genotype and phenotype information (63, 64). Although many pharmacogenetic laboratories utilize star nomenclature as recommended by PharmVar and CPIC, inter-laboratory differences in testing approaches and reporting remain. Clinical pharmacogenetic testing for *CYP2C19* can be performed on a variety of platforms using different methodologies and genotyping data can be reported in different ways, such as chromosomal or genomic position on reference sequence (RefSeq), amino acid change, dbSNP rsID, and/or using star (\*) allele nomenclature (12, 65, 66).

A recent study performed by the Genetic Testing Reference Material Program (GeT-RM) concluded that many pharmacogenetic variants interrogated were not consistent across platforms (67). Similar findings were reported by Moyer *et al.*, when they surveyed laboratories offering pharmacogenetic services for *CYP2D6* and *CYP2C19* genotyping (68). To address this, the Association for Molecular Pathology (AMP) and College of American Pathologists (CAP) have recently published recommendations for alleles to be included in clinical *CYP2C19* testing (24). In brief, the recommendations are based on allele function, allele frequencies across populations and ethnicities, and the availability of reference materials. The AMP working group recommended Tier 1 alleles (minimum panel of variant alleles), which includes *CYP2C19*\*2, \*3, and \*17, and Tier 2 (extended panel of variant alleles) that includes *CYP2C19*\*4.001 and \*4.002 (previously defined as \*4A and \*4B), \*5, \*6, \*7, \*8, \*9, \*10, and \*35. The utilization of star allele nomenclature as provided by PharmVar (14) will not only ensure that each star allele represents a unique and fully defined haplotype but also minimize “mis-interpretation” of a genotype result and its clinical implication(s).

The two end-user groups benefitting the most from standardized allele designations are clinicians and patients. Standardized terms and language will help clinician to convey and explain results and patients to understand what the results mean for them. Consistent nomenclature is a must for the integration of pharmacogenetics into electronic health records, as well as for the establishment of clinical decision support algorithms and the design of clinical support tools such as interruptive alerts (63, 64). For example, drug/allele combinations for alerts require systematic annotations within the EHR, using standardized nomenclature and terms. Finally, the analysis of pharmacogenetic clinical correlations will benefit from harmonized nomenclature of the gene variants, as well as support consistent test interpretation by providers across healthcare systems.

## The *CYP2C19* Gene Locus

The *CYP2C19* gene is composed of nine exons which encode a protein of 490 amino acids. The gene is located on chromosome 10q23.33 within a cluster of genes comprising the *CYP2C* gene family (8). The *CYP2C* gene locus contains four genes, *CYP2C8*, *CYP2C9*, *CYP2C18*, and *CYP2C19* (Figure 1) spanning a total of 486 kb. *CYP2C19* is the longest of the four genes with 93.9 kb of full-gene sequence. Due to high similarities between these genes, genotyping assay design requires particular attention to ensure specific amplification from the targeted gene to avoid false-positive test results. *CYP2C19*-genotyping assays are commercially available and commonly used for testing. However, this (and similar methods) require(s) the amplification of relatively short PCR fragments, which can make *CYP2C19*-specific primer design difficult for regions with high similarity with the other *CYP2C* genes, especially to *CYP2C9*.

Genetic variation, contributing to interindividual variability in *CYP2C19* activity, is caused by SNVs within coding regions. Gene copy number variation (CNVs) may also affect activity in rare cases. Numerous deletion and duplication events, affecting the *CYP2C* gene locus, have been described (69, 70). Many of the gene deletion and duplication events involve *CYP2C19*, as well as other *CYP2C* gene(s) and in some instances even many genes within the chromosomal region. Most pharmacogenetic genotyping tests typically do not interrogate for *CYP2C19* CNVs, as these variants are more commonly interrogated by copy number arrays and other quantitative molecular assays (e.g., qPCR, MLPA) (70). PharmVar has recently catalogued the *CYP2C19*\*36 and \*37 alleles, which represent full and partial (with including at least exon 1) *CYP2C19* gene deletions, respectively.

## *CYP2C19* allele, genotype, and phenotype frequencies across populations

The *CYP2C19* frequency table available at PharmGKB (44), 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, 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 (71). The *CYP2C19* 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, however, limitations regarding the accuracy of allele frequencies as follows: 1) frequencies are based on published allele data (limited for some populations), 2) most studies test for a limited number of allelic variants, which may lead to an underestimation of certain alleles. For example, c.-806C>T is often defaulted to a *CYP2C19*\*17 assignment, although this SNV is also present on *CYP2C19*\*4.002 (Figure 2). Likewise, if no SNVs are found,



*CYP2C19\*1* is assigned, which inflates the frequency of this allele; 3) limited data for CNVs, and 4) errors translating SNV which consequently may lead to over- or under-reporting of certain alleles or some alleles not being detected at all. Therefore, all calculations based on allele frequencies are estimates at best and should be used with caution.

Regardless, there is considerable variation among the estimated frequencies for individual alleles across biogeographical groups. The no function *CYP2C19\*2* allele has been found at high frequencies in Oceanian (60%) and South and East Asian (28%) populations. This allele is less frequent, albeit still common, in Sub-Saharan African populations and their descendants (16–18%), and Europeans (15%). *CYP2C19\*3*, also a no function allele, is more frequent in Oceanians (15%), compared to East Asians (7%) and Near Eastern populations (2%), and is rarely found in other populations around the world (e.g., 0.3% in African-American, Afro-Caribbean and Sub-Saharan Africans and 0.2% in Europeans). Other alleles, including *CYP2C19\*9* (decreased function), *\*15* (normal function), and *\*35* (no function), are more prevalent in African populations. In contrast, the *CYP2C19\*17* increased function allele is frequently observed in Europeans (22%), Near Eastern populations (19%), Sub-Saharan African (17%), but rarely in East Asians (2%). Consequently, the highest frequency of CYP2C19 PMs are found in Oceanic and East Asian populations. Also, the frequency of genetic variants as well as their distribution and combination within a haplotype may differ in admixed populations when compared to parental populations. For example, *CYP2C19* has a higher number of rare genetic variants in Hispanics (4%), a highly admixed population, when compared to Europeans (0.05%) and *CYP2C19\*2* and *\*17* were found at a higher frequency in self-reported Afro-Caribbean subjects from Costa Rica than mestizos of the same country (72, 73). In addition, frequencies may also profoundly differ among indigenous populations including those in North, South, and Central America (74-76).

Considering *CYP2C19\*1* through *\*35* (excluding the *CYP2C19\*36* and *\*37* deletion alleles), hundreds of allele combinations are possible, and thus, the number of genotypes in a population or patient cohort can be quite large. The actual number of combinations that occur in each population may be significantly less, however, 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 CYP2C19 Frequency Table (44). We stress, however, to view all phenotype group frequencies (including those shown in the PharmGKB/CPIC table) 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’ (77).

## CYP2C19 Allele function

PharmVar displays CPIC allele clinical function, using respective CPIC terms (increased, normal, decreased or no function or uncertain or unknown function. Alleles that have not been assigned function by CPIC are shown ‘not available’. The filter option allows the user to sort alleles by functional status. See sections above regarding details of CPIC function assignment.

## PharmVar Nomenclature and *CYP2C19* allele designation

PharmVar stores and displays allelic data consistently across genes, relying on public standards and data sources wherever possible. The standardized nomenclature follows criteria developed by gene experts. The “Allele Designation and Evidence Level Criteria” document describes the nomenclature system and provides examples (78). For example, a new star number is only issued if a haplotype contains a sequence variant that: 1) results in an amino acid change, e.g., *CYP2C19*\*5 harbors an arginine to tryptophan change (c.1297C>T, p.R433W), 2) abolishes a splice site (e.g., c.681G>A in exon 5 of *CYP2C19*\*2 alters a splice site that leads to a frameshift and premature translation termination), or 3) changes expression levels causing decreased or increased function, which is exemplified by the *CYP2C19*\*17 defining variant (c.-806C>T) that increases expression. Importantly, new haplotypes that contain previously characterized variants that obliterate function are catalogued under the original star (\*) allele number as a suballele. For example, any allele having a novel variant and c.1297C>T (p.R433W), will be designated as a *CYP2C19*\*5 suballele and considered having no function, regardless of the functional status of the novel variant.

### The PharmVar *CYP2C19* gene expert panel

International experts representing *CYP2C19* research, clinical testing, and implementation interests were recruited from PharmVar members to serve on the *CYP2C19* expert panel and tasked with reviewing the current nomenclature and any new submissions. The panel also includes PharmGKB/CPIC representation to ensure that the nomenclature is consistent with CPIC guidelines and to facilitate dissemination to a greater audience through PharmGKB as well as other databases, such as ClinGen. The *CYP2C19* expert panel first convened in January 2018 and meets monthly via teleconferences and communicates regularly by email as needed. The composition of the panel can be found on the PharmVar website (79). Table 2 summarizes the alleles received and accepted by the panel.

### The PharmVar *CYP2C19* gene page

PharmVar maps sequence variations for each gene to its latest genomic and cDNA reference sequences (RefSeqs), issued by the NCBI Reference Sequences database (80), and the GRCh37 (NC\_000010.10) and GRCh38 (NC\_000010.11) genome builds. For *CYP2C19*, the current genomic and transcript RefSeqs are NG\_008384.3 and NM\_000769.1, respectively. *CYP2C19* was transitioned from the original nomenclature site to PharmVar in Sept 26, 2017 (original content is available through the ‘Archive’ link on the PharmVar homepage).

It is important to note that GRCh37 differs from GRCh38 in three positions, 96521422A>G (c.-1041G>A), 96522561T>C (c.99C>T) and 96602623G>A (c.991A>G). Therefore, the number of variants displayed for GRCh37 differs from those shown for GRCh38, which corresponds to the variant that was originally designated as *CYP2C19*\*27 and was reclassified as \*1.006.

A Locus Reference Genomic (LRG) record was requested by PharmVar from the LRG Project, a NCBI (RefSeq) and EMBL-EBI (Ensembl/GENCODE) initiative (81). 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 after their release. The LRG for *CYP2C19* (LRG\_584, released 6–3-2020) matches 100% with the NG\_008384.3 RefSeq and will be used by PharmVar as the ‘gold-standard’ reference moving forward.

PharmVar is left-aligning, i.e. insertions and deletions of nucleotides in a repeat or homopolymer sequence are listed using the 3’ rule recommended by HGVS (see PharmVar ‘Standards’ for more details (82) or the *CYP2D6* GeneFocus (61) for examples). Due to different alignment methods, coordinates for insertion/deletion polymorphisms may differ among databases. No such variants have, however, been described for *CYP2C19* to date.

On the PharmVar *CYP2C19* gene page, the user can easily cross-reference genomic and cDNA position(s) by choosing the respective reference sequence or genome build of interest; there is also the option of two count modes, i.e. counting from the first nucleotide in the reference sequence or the ATG translation start codon being +1.

## **CYP2C19 haplotype evidence levels**

PharmVar designates the “Haplotype Evidence Level” for each of the star alleles reported on the *CYP2C19* gene page. Evidence levels are displayed as symbols indicating ‘definitive’ (Def), ‘moderate’ (Mod) or ‘limited’ (Lim) levels of support for a given haplotype. This three-tiered system represents a modified ClinVar classification system; more detailed information is provided in the ‘Allele Designation and Evidence Level’ document (78). This type of information (e.g., whether an allele was sequenced across the gene, how a haplotype was determined) was not always systematically captured prior to PharmVar. For existing haplotype definitions, a literature review was conducted in order to assign evidence levels. Many alleles are currently labeled as ‘Lim’ because their definitions do not include any, or the complete 2 kb of upstream sequence required for each submission (as set by the *CYP2C* expert panel). This was the case for many allelic variants including *CYP2C19*\*2.001 and \*3.001. Other alleles were labeled as ‘Mod’ although the required upstream region was sequenced, but the phase of sequence variants was computationally inferred and has not been validated (e.g., *CYP2C19*\*9-\*14). The value of evidence levels is centered on providing users with as much information on haplotype reliability as possible and enabling users to quickly parse haplotypes based on robust, high evidence, versus other haplotypes with ‘Lim’ or ‘Mod’ evidence levels.

PharmVar solicits submissions for all alleles labeled ‘Lim’ and ‘Mod’ to ultimately raise their evidence levels to ‘Def’. Moreover, PharmVar also encourages encore submissions for alleles with single citations and shown as ‘Def’ to further corroborate a haplotype definition.

## **PharmVar IDs**

Each characterized haplotype receives a PharmVar ID (PVID). The PVID is a unique numeric identifier analogous to dbSNP rsIDs. Star allele names are driven by functional

grouping, i.e., they are not guaranteed to be permanent and can be subject to change (for details see “corrections revisions, new alleles, updates” section below). Additional changes may be necessary in the future as more information becomes available. If an allele’s star designation is updated to a new star number, the PVID of the haplotype remains constant and does not change. In contrast, if a haplotype definition changes (e.g., through the addition or removal of variants) a new PVID will be assigned. An example is the above mentioned *CYP2C19\*27* allele. Its initial PVID (PV00073) remained the same. In contrast, *CYP2C19\*28* received a new PVID (PV00622) after its haplotype was revised (i.e. two SNVs were removed). Original PVIDs and their haplotype definitions can be tracked in the database via the PVID Lookup 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:

Haplotype definitions for *CYP2C19* allelic variation is based on the sequence of the coding regions (counting from the start and stop codons, respectively). Partial or incomplete gene sequences are not accepted by PharmVar. Many of the allelic variants that have been transitioned from the original nomenclature site to PharmVar have limited information, such as: 1) upstream and downstream regions were not, or only partially sequenced, 2) exons were not sequenced, but analyzed through different methodologies, 3) the haplotype definition was inferred and therefore may have SNVs in exons, flanking intronic, and/or up- and downstream regions that have not been detected.

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

Coordinates were mapped to NG\_008384.2 by the nomenclature site and updated to the current *CYP2C19* RefSeq NG\_008384.3 as the gene was transitioned to PharmVar. This upgrade did not change position annotations of SNVs.

As *CYP2C19* was transitioned into the PharmVar database, comments and footnotes were removed and errors corrected. References in support of allele definitions have been updated and those solely describing function removed (references for function are provided in the PharmGKB/CPIC *CYP2C19* Allele Functionality table (44). A number of descriptors such as “g.12662A>G is likely part of all \*2 alleles” have been removed and the intronic SNV at g.12662A>G (c.332–23A>G), which also causes a splicing defect, has been added to all *CYP2C19\*2* haplotype definitions. It remains unknown, however, whether all *CYP2C19\*2* haplotypes indeed have this SNV in addition to c.681G>A, which is shared among all *CYP2C19\*2* alleles. Changes and revisions are detailed in the ‘Change Log’ document on the *CYP2C19* gene page (83).

The expert panel recommended to reassign *CYP2C19\*27* as a *CYP2C19\*1* suballele (*CYP2C19\*1.006*), due to limited information regarding c.–1041G>A. This allele was

originally given its own star designation based on *in vitro* data, suggesting that c.-1041G>A may decrease expression levels (84). There is no *in vivo* evidence, however, supporting the initial findings.

As of February 2020, the *CYP2C19* expert panel has designated two novel alleles (\*36 and \*37) and three new suballeles. In addition, five alleles that were based on partial information are now fulfilling PharmVar allele definition requirements and their evidence level was raised to 'Def' (Table 2). There are still numerous allele definitions with 'Lim' and 'Mod' evidence levels for which PharmVar seeks submissions. The 'Change Log' document also tracks submissions and indicates the star alleles that have been updated.

### Challenges of allele designation:

Despite detailed allele definition criteria, assigning star allele definitions can be challenging as illustrated by two independent submissions. The haplotype of the first submission was designated as a *CYP2C19*\*1.005 suballele. The second submission, accompanied by unpublished *in vitro* data, was by the author who previously published this haplotype (85), requesting that this allele receive its own star number. Upon careful review and communication with the submitter, the panel unanimously agreed that the published and unpublished data were inconsistent and not sufficient for issuing this allelic variant its own star name.

Upon reviewing the limited data available for *CYP2C19*\*27, the panel recommended to reclassify this allelic variant as a *CYP2C19*\*1 suballele (noting that this haplotype should not have received a new star number when it was first accepted by the P450 Nomenclature Committee, given the limited *in vitro* functional data). The re-assignment of *CYP2C19*\*27 as *CYP2C19*\*1.006 is consistent with the PharmVar criteria, as well as consistent of how the criteria were applied to the allele designated *CYP2C19*\*1.005. These examples illustrate that allele designation is not always straightforward and needs to be a dynamic process to accommodate new findings.

### Update and reassignment of *CYP2C19*\*1 (*CYP2C19*\*1.001):

Since the inception of PharmVar, the *CYP2C19* panel and Steering Committee discussed that the previously issued *CYP2C19*\*1 suballele definitions do not conform to formal PharmVar allele designation criteria (78). The *CYP2C19*\*1.001 lacks c.991A>G (p.I331V) while this amino acid changing SNV is present in the *CYP2C19*\*1.002-\*1.006 suballeles. The original designation of the latter as *CYP2C19*\*1 suballeles was most likely due to the interpretation of p.I331V having no effect on enzyme activity. The common c.991A>G (p.I331V) variant is also found on many other *CYP2C19* haplotypes and thus represents the major allele across most populations (c.991A>G, G=0.9386 per the gnomAD exome database). This prompted the question of whether the haplotype representing *CYP2C19*\*1.002 would be a more appropriate reference sequence for allele definition (i.e., if allele definitions would be reported against *CYP2C19*\*1.002 as reference, the majority of allelic variants would not have c.991A>G, nor c.99C>T, a synonymous SNV that is present on the majority of alleles as well). Considering a current massive standardization effort undertaken by the NCBI and EMBL-EBI, neither the *CYP2C19* RefSeq, nor the LRG will

be revised to accommodate star nomenclature. Based on a majority consensus, the allele originally catalogued as *CYP2C19\*1.001* (i.e., *\*1A*) has now received its own star designation, i.e., *CYP2C19\*38*. The core allele definition of this variant does not have any SNVs because this haplotype matches the NG\_008384.3 RefSeq. In contrast, the *CYP2C19\*1* allele definition, representing the most frequently observed *CYP2C19* allele, gained a core SNP because all variants catalogued under this designation have p.I331V.

RefSeqs are generally representing a haplotype that encodes functional protein and is commonly found across populations, this however is not always the case (the *CYP3A5* RefSeq, for example, represents functional protein, but the non-functional *CYP3A5\*3* variant is the most common across many populations). Also, a RefSeq may have changed over the past two decades, as it has for *CYP2C19* (the genomic sequences among GRCh37 and GRCh38 differ in three positions, i.e. c.-1041, c.99 and c.991) making consistent reporting of alleles difficult. As the NCBI is standardizing reference sequences and releases LRGs, PharmVar has not only a unique opportunity, but also an obligation to follow suit, and align allele definitions with globally accepted RefSeqs as well as update definitions that are inconsistent. Although changes such as the *CYP2C19\*1.001* to *\*38* update may be viewed as ‘interruptive’, having a standardized nomenclature system moving forward is preferred as reflected by the expert consensus vote.

What does this update mean for the interpretation of genotyping test results? If no SNVs are found, and c.991A>G (p.I331V) is not tested, a *CYP2C19\*1* will be assigned based on defaulting to the more common allele having c.991G. Clinical genotyping tests typically do not include c.991A>G (p.I331V) (it is also not recommended by AMP for Tier 1 or 2 testing (24)) and thus *CYP2C19\*38* alleles will be called as *\*1* by default. If genotype is based on sequencing and information for c.991A>G (p.I331V) is available, a small fraction of alleles will be called as *CYP2C19\*38*. The introduction of *CYP2C19\*38* will not impact phenotype assignments as it will be designated as having normal function.

## Core allele definitions

For many alleles there is a growing number of suballeles that share one or more ‘key’ defining sequence variant(s), which are considered ‘core’ SNVs by PharmVar. Although suballele information can be valuable for the design of test platforms (sequence or genotype-based alike) and interpretation of genotyping test results, there is no need to distinguish suballeles for phenotype prediction because all alleles under a star number are defined per criteria as functionally equivalent. Thus, even if a test is capable of distinguishing suballeles, these can be simply reported as the defining allele (e.g., *CYP2C19\*1*, *\*2*, etc.)

PharmVar and the PharmGKB have collaboratively developed core allele definitions. A core allele is defined only by sequence variations that cause an amino acid change or impact function by changing expression levels or interfere with splicing and are present in all suballeles within an allele group. With this rule-based system, suballeles are collapsed into a single ‘core’ definition representing all suballeles categorized under a star (\*) number. For example, the growing number of *CYP2C19\*2* suballeles share two SNVs that fulfill this rule, i.e. c.332–23A>G (g.12662A>G) and c.681G>A (g.19154G>A). Both SNVs have been

shown to cause aberrant splicing, preventing the formation of a functional gene product. Thus, both SNVs constitute the *CYP2C19\*2* core allele definition (Figure 3). For *CYP2C19\*4*, only c.1A>G (p.M1V) that abolishes translation initiation is shared among all subvariants and therefore is the sole variant of the *\*4* core allele definition (Figure 2).

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 *CYP2C19\*2* core definition (Figure 2). Indeed, c.332–23A>G is also present in *CYP2C19\*35* and constitutes the sole core SNV for this allele. Another example is c.–806C>T which increases mRNA expression and hence, fulfills core SNV definition requirements for *CYP2C19\*17*. This SNV is, however, also part of one of the two known *CYP2C19\*4* haplotypes.

One challenge with core allele definitions is that a definition may change over time as new information becomes available. For instance, based on the rules described above, the current core allele definition of *CYP2C19\*2* harbors c.332–23A>G, c.681G>A, and c.990C>T (p.I331V). According to the *CYP2C19* allele designation criteria, all new alleles having c.681G>A, and therefore are nonfunctional due to aberrant splicing, will be assigned as *CYP2C19\*2* suballeles, regardless of the nature of the other SNV(s) present. Consequently, the *CYP2C19\*2* core allele definition may change, should a sequence containing c.681G>A, but not c.332–23A>G, and/or c.990C>T (p.I331V) be identified.

The core alleles are the basis of the *CYP2C19* allele definition table used in CPIC guidelines and by PharmGKB (this table is available to PharmGKB users as definition material (44)). The *CYP2C19* core allele definitions are also utilized for clinical annotations in PharmGKB.

## The PharmVar Comparative Allele ViewEr

The *Comparative Allele ViewEr* (CAVE) tool was developed by PharmVar to easily compare core alleles (86). This tool can be accessed using the “Compare View” button on the *CYP2C19* gene page. Figure 2C exemplifies the utility of this tool on *CYP2C19\*2*, *\*4*, *\*17* and *\*35*. In this display mode it is easy to see which core SNVs are shared among haplotypes, whether they alter function and/or are unique to a haplotype.

## Reporting genotype and translation into phenotype

PharmVar and PharmGKB have also collaboratively developed templates to facilitate more consistent and transparent reporting of genotype details and how genotype is translated into phenotype to be used by the community to include more detailed information as part of the submission of research findings to publishers. 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 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 was translated into phenotype in a study, as well as genotype frequencies. Although it is recommended by CPIC, as well as other groups, to use their standardized translation method, not every investigator or laboratory adopts this method. Too often,

papers reference previous work stating that ‘genotyping was performed as previously described’ or indicate that ‘CYP2C19 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, for curators to compare findings or extract 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.

### **CYP2C19 reference materials**

The GeT-RM is a combined effort among the Centers for Disease Control and Prevention–based Genetic Testing Reference Material Coordination Program, Coriell Institute for Medical Research, and members of the pharmacogenetic testing community. Considering the growing use of pharmacogenetic testing, established sets of well-characterized reference materials are needed for assay development, validation, quality control, and proficiency testing. To address the increasing need for reference materials, a set of 137 genomic DNA samples were characterized for 28 pharmacogenes, including *CYP2C19* and “consensus” genotypes established (67). Although the most common variants were assayed, many rare alleles were not identified among the samples tested; efforts are ongoing to find additional samples to complement the existing materials for *CYP2C19*. Testing and research laboratories can acquire these materials from the Coriell Institute (Camden, NJ, USA), as they are publicly available.

### **Inferring CYP2C19 haplotype from Next Generation Sequence data and public databases**

Bioinformatic tools have been developed to call star (\*) alleles from next-generation sequencing (NGS) panels and/or exome/genome-based sequence data for several *CYP* genes. Available tools include Aldy (87), Stargazer (88), Astrolabe (89, 90), and PharmCAT (36). These tools use algorithms to predict the individual diplotype, based on the known alleles catalogued by PharmVar. These programs have not yet been systematically validated, however. In addition, high quality data are needed in order to generate reliable genotype calls with NGS. Finally, CNVs affecting the *CYP2C19* gene locus may also lead to misinterpretation in rare cases (70).

### **Methods for CYP2C19 allele characterization**

As mentioned above, the *CYP2C19* gene spans almost 93 kb, which makes haplotype analysis non-trivial, especially in cases when the SNVs for which linkage needs to be established exceed the distance that can be covered by XL-PCR amplification (typically under 20 kb). If a subject is heterozygous for a single SNV or homozygous for one or more SNVs, the haplotype is unambiguous. If a novel SNV (or SNVs) is detected along a known SNV (or SNVs), of which all are heterozygous, it remains unknown, whether the novel SNV(s) is/are in *cis* or *trans* with the other SNV(s). Function may depend on which SNVs are occurring together on the same chromosome; the latter will also dictate whether the novel haplotype will be cataloged as a suballele or receive its own star number. Haplotypes



may also be computationally inferred, which may, however, not result in high-confidence predictions, especially if the novel haplotype was observed in a single individual.

Thus, the gold standard for allele definition is experimental validation. Figure 4 provides an overview of methods/approaches to characterize a novel haplotype. Figure 4A demonstrates that pedigree analysis can be used to determine, whether a novel SNV is in *cis* or *trans* with those defining the *CYP2C19\*2* allele. Figure 4B illustrates, that allele-specific PCR employing primers, that only amplify, when SNVs are either in *cis* or *trans*, can be utilized to establish SNV linkage; this approach, however, only works, if the SNVs to be linked, are within a distance that can reliably be amplified by PCR. Figure 4C illustrates the utility of single molecule real time sequencing methods (e.g., Pacific Biosciences (Menlo Park, CA) or Oxford Nanopore Technologies (Oxford, United Kingdom) and 10X Genomics Long-Read mapping (Pleasanton, CA)) to establish haplotypes over a range of distances. SNVs can also be phased over long distances of up to 200 kb using digital droplet PCR (ddPCR) to establish haplotypes (91, 92). In Figure 4D we offer an example of a PharmVar submission for *CYP2C19\*35*. The haplotype of the *CYP2C19\*35* was determined from whole genome sequence data of each member of the family trio (father, mother and child). The father is homozygous for the *CYP2C19\*17.001* allele and thus passes these to his offspring. All SNVs encompassing the child's *CYP2C19\*35* allele must have been inherited from the mother. This analysis revealed that the allele has an additional SNV, c.-913G>A, which was not part of the initial haplotype definition. Based on this information, the *CYP2C19\*35* allele definition was updated, and its evidence level revised from 'Lim' to 'Def'.

To detect CNVs, methods such as multiplex ligation-dependent probe amplification (MLPA), TaqMan® copy number assays or ddPCR may be employed to evaluate or confirm *CYP2C19* gene copy number status. *CYP2C19* CNVs are rare and are most often detected as incidental findings by array platforms such as array comparative genomic hybridization (aCGH), performed for unrelated diagnostic purposes.

## Conclusions

This is the second of a series of gene-centric review articles, focusing on important pharmacogenes. The *CYP2C19* GeneFocus provides essential information for the understanding of this highly polymorphic gene, complementing clinically relevant information provided by CPIC guidelines and other pharmacogenetic resources. We are highlighting PharmVar efforts and challenges of systematically cataloging *CYP2C19* 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

We would like thank Mia Wadelius, Uppsala University, Sweden, for serving on the PharmVar *CYP2C* expert panel since its inception and critical comments on the manuscript. We also like to sincerely thank Roger Gaedigk for his assistance with graphical artwork and critically reading the manuscript.

## Funding

This work was funded by the National Institutes of Health for the Pharmacogene Variation Consortium (R24 GM123930; PI, A.G.) and PharmGKB (U24 HG010615; PI, T.E.K.). V.M.P. is supported by the Implementing Genomics in Practice (IGNITE) project grant (U01 HG010245). J.D. is supported by grant U54 MD007600–33 (RCMI Program) from NIH/NIMHD.

## References

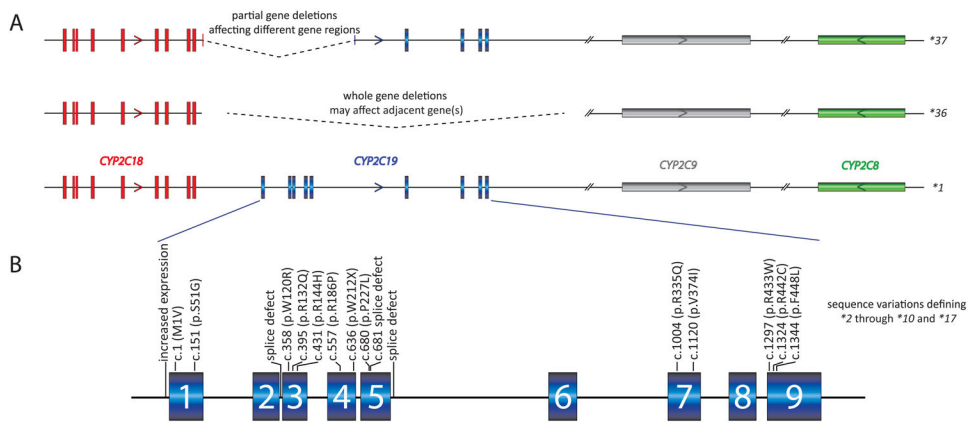
- (1). Kupfer A & Preisig R Pharmacogenetics of mephenytoin: a new drug hydroxylation polymorphism in man. *Eur J Clin Pharmacol* 26, 753–9 (1984). [PubMed: 6489416]
- (2). Goldstein JA et al. Evidence that CYP2C19 is the major (S)-mephenytoin 4'-hydroxylase in humans. *Biochemistry* 33, 1743–52 (1994). [PubMed: 8110777]
- (3). Wrighton SA, Stevens JC, Becker GW & VandenBranden M Isolation and characterization of human liver cytochrome P450 2C19: correlation between 2C19 and S-mephenytoin 4'-hydroxylation. *Arch Biochem Biophys* 306, 240–5 (1993). [PubMed: 8215410]
- (4). Romkes M, Faletto MB, Blaisdell JA, Raucy JL & Goldstein JA Cloning and expression of complementary DNAs for multiple members of the human cytochrome P450IIC subfamily. *Biochemistry* 30, 3247–55 (1991). [PubMed: 2009263]
- (5). Goldstein JA & de Morais SM *Biochemistry and molecular biology of the human CYP2C subfamily. Pharmacogenetics* 4, 285–99 (1994). [PubMed: 7704034]
- (6). de Morais SM, Wilkinson GR, Blaisdell J, Nakamura K, Meyer UA & Goldstein JA The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans. *J Biol Chem* 269, 15419–22 (1994). [PubMed: 8195181]
- (7). Meyer UA & Zanger UM Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annu Rev Pharmacol Toxicol* 37, 269–96 (1997). [PubMed: 9131254]
- (8). Scott SA et al. PharmGKB summary: very important pharmacogene information for cytochrome P450, family 2, subfamily C, polypeptide 19. *Pharmacogenet Genomics* 22, 159–65 (2012). [PubMed: 22027650]
- (9). Sim SC et al. A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. *Clin Pharmacol Ther* 79, 103–13 (2006). [PubMed: 16413245]
- (10). 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]
- (11). Daly AK et al. Nomenclature for human CYP2D6 alleles. *Pharmacogenetics* 6, 193–201 (1996). [PubMed: 8807658]
- (12). 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]
- (13). Gaedigk A et al. The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clin Pharmacol Ther* 103, 399–401 (2018). [PubMed: 29134625]
- (14). Gaedigk A, Whirl-Carrillo M, Pratt VM, Miller NA & Klein TE PharmVar and the Landscape of Pharmacogenetic Resources. *Clin Pharmacol Ther* 107, 43–6 (2020). [PubMed: 31758698]
- (15). PharmGKB CYP2C19 Drug Labels Annotations. Accessed 06 April 2020
- (16). FDA Table of Pharmacogenomics Biomarkers in Drug Labeling <https://www.fda.gov/drugs/science-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling>. Accessed 06 April 2020.
- (17). FDA Table of Pharmacogenetics Associations <https://www.fda.gov/medical-devices/precision-medicine/table-pharmacogenetic-associations>. Accessed 06 April 2020.
- (18). CPIC Gene-drug pairs <https://cpicpgx.org/genes-drugs>. Accessed 06 April 2020.
- (19). Cavallari LH et al. The IGNITE Pharmacogenetics Working Group: An Opportunity for Building Evidence with Pharmacogenetic Implementation in a Real-World Setting. *Clin Transl Sci* 10, 143–6 (2017). [PubMed: 28294551]

- (20). Cavallari LH et al. Multisite Investigation of Outcomes With Implementation of CYP2C19 Genotype-Guided Antiplatelet Therapy After Percutaneous Coronary Intervention. *JACC Cardiovasc Interv* 11, 181–91 (2018). [PubMed: 29102571]
- (21). Claassens DMF et al. A Genotype-Guided Strategy for Oral P2Y12 Inhibitors in Primary PCI. *N Engl J Med* 381, 1621–31 (2019). [PubMed: 31479209]
- (22). Bergmeijer TO et al. Genome-wide and candidate gene approaches of clopidogrel efficacy using pharmacodynamic and clinical end points-Rationale and design of the International Clopidogrel Pharmacogenomics Consortium (ICPC). *Am Heart J* 198, 152–9 (2018). [PubMed: 29653637]
- (23). Patel JN et al. Evaluation of CYP2C19 Genotype-Guided Voriconazole Prophylaxis After Allogeneic Hematopoietic Cell Transplant. *Clin Pharmacol Ther* 107, 571–9 (2020). [PubMed: 31549386]
- (24). 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]
- (25). 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]
- (26). 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]
- (27). Malki MA & Pearson ER Drug-drug-gene interactions and adverse drug reactions. *Pharmacogenomics J*, 20, 355–366 (2020). [PubMed: 31792369]
- (28). Zhou ZW et al. Clinical association between pharmacogenomics and adverse drug reactions. *Drugs* 75, 589–631 (2015). [PubMed: 25895462]
- (29). Maher RL, Hanlon J & Hajjar ER Clinical consequences of polypharmacy in elderly. *Expert Opin Drug Saf* 13, 57–65 (2014). [PubMed: 24073682]
- (30). Drug Interactions Flockhart TableTM. <https://edswi.org/wp-content/uploads/2018/02/Flockhart-Table-Medication-Metabolism.pdf>. Accessed 06 April 2020.
- (31). Gravel S, Chiasson JL, Turgeon J, Grangeon A & Michaud V Modulation of CYP450 Activities in Patients With Type 2 Diabetes. *Clin Pharmacol Ther* 106, 1280–9 (2019). [PubMed: 31099895]
- (32). Angiolillo DJ et al. Randomized comparison of a high clopidogrel maintenance dose in patients with diabetes mellitus and coronary artery disease: results of the Optimizing Antiplatelet Therapy in Diabetes Mellitus (OPTIMUS) study. *Circulation* 115, 708–16 (2007). [PubMed: 17261652]
- (33). Zanger UM & Schwab M Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* 138, 103–41 (2013). [PubMed: 23333322]
- (34). Laine K, Tybring G & Bertilsson L No sex-related differences but significant inhibition by oral contraceptives of CYP2C19 activity as measured by the probe drugs mephenytoin and omeprazole in healthy Swedish white subjects. *Clin Pharmacol Ther* 68, 151–9 (2000). [PubMed: 10976546]
- (35). Hagg S, Spigset O & Dahlqvist R Influence of gender and oral contraceptives on CYP2D6 and CYP2C19 activity in healthy volunteers. *Br J Clin Pharmacol* 51, 169–73 (2001). [PubMed: 11259990]
- (36). Xie HG, Huang SL, Xu ZH, Xiao ZS, He N & Zhou HH Evidence for the effect of gender on activity of (S)-mephenytoin 4'-hydroxylase (CYP2C19) in a Chinese population. *Pharmacogenetics* 7, 115–9 (1997). [PubMed: 9170148]
- (37). Svirni S, Shpizen S, Leitersdorf E, Levy M & Caraco Y Phenotypic-genotypic analysis of CYP2C19 in the Jewish Israeli population. *Clin Pharmacol Ther* 65, 275–82 (1999). [PubMed: 10096259]
- (38). Yanni SB, Annaert PP, Augustijns P, Ibrahim JG, Benjamin DK Jr. & Thakker DR In vitro hepatic metabolism explains higher clearance of voriconazole in children versus adults: role of CYP2C19 and flavin-containing monooxygenase 3. *Drug Metab Dispos* 38, 25–31 (2010). [PubMed: 19841059]

- (39). Zane NR, Chen Y, Wang MZ & Thakker DR Cytochrome P450 and flavin-containing monooxygenase families: age-dependent differences in expression and functional activity. *Pediatr Res* 83, 527–35 (2018). [PubMed: 28922349]
- (40). Koukouritaki SB et al. Developmental expression of human hepatic CYP2C9 and CYP2C19. *J Pharmacol Exp Ther* 308, 965–74 (2004). [PubMed: 14634042]
- (41). Burgess KS et al. Age-Related Changes in MicroRNA Expression and Pharmacogenes in Human Liver. *Clin Pharmacol Ther* 98, 205–15 (2015). [PubMed: 25968989]
- (42). Whirl-Carrillo M et al. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* 92, 414–7 (2012). [PubMed: 22992668]
- (43). PharmGKB CYP2C19 gene page <https://www.pharmgkb.org/gene/PA124>. Accessed 06 April 2020.
- (44). PharmGKB Gene-specific Information Tables for CYP2C19 <https://www.pharmgkb.org/page/cyp2c19RefMaterials>. Accessed 06 April 2020.
- (45). Sangkuhl K et al. Pharmacogenomics Clinical Annotation Tool (PharmCAT). *Clin Pharmacol Ther* 107, 203–10 (2020). [PubMed: 31306493]
- (46). 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]
- (47). Relling MV, Klein TE, Gammal RS, Whirl-Carrillo M, Hoffman JM & Caudle KE The Clinical Pharmacogenetics Implementation Consortium: 10 Years Later. *Clin Pharmacol Ther* 107, 171–5 (2020). [PubMed: 31562822]
- (48). 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]
- (49). 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]
- (50). Scott SA et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin Pharmacol Ther* 94, 317–23 (2013). [PubMed: 23698643]
- (51). Moriyama B et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP2C19 and Voriconazole Therapy. *Clin Pharmacol Ther* 102, 45–51 (2017). [PubMed: 27981572]
- (52). 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]
- (53). Caudle KE et al. Standardizing CYP2D6 Genotype to Phenotype Translation: Consensus Recommendations from the Clinical Pharmacogenetics Implementation Consortium and Dutch Pharmacogenetics Working Group. *Clin Transl Sci* 13, 116–24 (2020). [PubMed: 31647186]
- (54). Xu RA, Gu EM, Liu TH, Ou-Yang QG, Hu GX & Cai JP The effects of cytochrome P450 2C19 polymorphism on the metabolism of voriconazole in vitro. *Infect Drug Resist* 11, 2129–35 (2018). [PubMed: 30464555]
- (55). Wang H et al. Evaluation of the effects of 20 nonsynonymous single nucleotide polymorphisms of CYP2C19 on S-mephenytoin 4'-hydroxylation and omeprazole 5'-hydroxylation. *Drug Metab Dispos* 39, 830–7 (2011). [PubMed: 21325430]
- (56). Takahashi M et al. Functional characterization of 21 CYP2C19 allelic variants for clopidogrel 2-oxidation. *Pharmacogenomics J* 15, 26–32 (2015). [PubMed: 25001882]
- (57). Dai DP et al. In vitro functional analysis of 24 novel CYP2C19 variants recently found in the Chinese Han population. *Xenobiotica* 45, 1030–5 (2015). [PubMed: 26153442]
- (58). Lau PS, Leong KV, Ong CE, Dong AN & Pan Y In Vitro Functional Characterisation of Cytochrome P450 (CYP) 2C19 Allelic Variants CYP2C19\*23 and CYP2C19\*24. *Biochem Genet* 55, 48–62 (2017). [PubMed: 27578295]

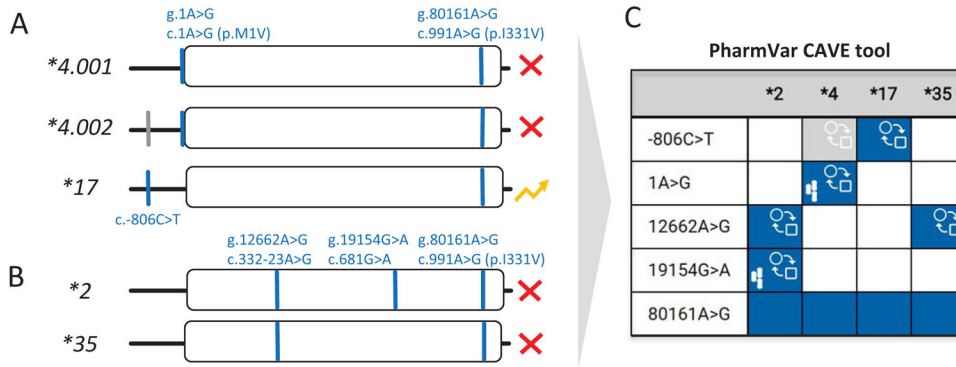
- (59). Zhang H, De T, Zhong Y & Perera MA The advantages and challenges of diversity in Pharmacogenomics: Can minority populations bring us closer to implementation? *Clin Pharmacol Ther* 106, 338–49 (2019). [PubMed: 31038731]
- (60). Sirugo G, Williams SM & Tishkoff SA The Missing Diversity in Human Genetic Studies. *Cell* 177, 1080 (2019). [PubMed: 31051100]
- (61). Nofziger C et al. PharmVar GeneFocus: CYP2D6. *Clin Pharmacol Ther* 107, 154–70 (2020). [PubMed: 31544239]
- (62). Cavallari LH et al. Multi-site investigation of strategies for the clinical implementation of CYP2D6 genotyping to guide drug prescribing. *Genet Med* 21, 2255–63 (2019). [PubMed: 30894703]
- (63). 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]
- (64). Hicks JK, Dunnenberger HM, Gumpfer KF, 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). Gray KA, Yates B, Seal RL, Wright MW & Bruford EA [Genenames.org](https://www.genenames.org/): the HGNC resources in 2015. *Nucleic Acids Res* 43, D1079–85 (2015). [PubMed: 25361968]
- (66). The NCBI Handbook - *MedGen*. (2018).
- (67). 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]
- (68). Moyer AM, Rohrer Vitek CR, Giri J & Caraballo PJ Challenges in Ordering and Interpreting Pharmacogenomic Tests in Clinical Practice. *Am J Med* 130, 1342–4 (2017). [PubMed: 28757317]
- (69). Santos M et al. Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genet Med* 20, 622–9 (2018). [PubMed: 29261188]
- (70). Botton MR et al. Structural variation at the CYP2C locus: Characterization of deletion and duplication alleles. *Hum Mutat* 40, e37–e51 (2019). [PubMed: 31260137]
- (71). Huddart R et al. Standardized Biogeographic Grouping System for Annotating Populations in Pharmacogenetic Research. *Clin Pharmacol Ther* 105, 1256–62 (2019). [PubMed: 30506572]
- (72). Cespedes-Garro C et al. Relevance of the ancestry for the variability of the Drug-Metabolizing Enzymes CYP2C9, CYP2C19 and CYP2D6 polymorphisms in a multiethnic Costa Rican population. *Rev Biol Trop* 64, 1067–76 (2016). [PubMed: 29461783]
- (73). Zhou Y, Ingelman-Sundberg M & Lauschke VM Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. *Clin Pharmacol Ther* 102, 688–700 (2017). [PubMed: 28378927]
- (74). Roco A et al. Frequencies of 23 functionally significant variant alleles related with metabolism of antineoplastic drugs in the Chilean population: comparison with Caucasian and Asian populations. *Front Genet* 3, 229 (2012).doi: 10.3389/fgene.2012.00229 [PubMed: 23130019]
- (75). Henderson LM et al. P450 Pharmacogenetics in Indigenous North American Populations. *J Pers Med* 8, 9 (2018). doi: 10.3390/jpm8010009.
- (76). Salazar-Flores J et al. Distribution of CYP2D6 and CYP2C19 polymorphisms associated with poor metabolizer phenotype in five Amerindian groups and western Mestizos from Mexico. *Genet Test Mol Biomarkers* 16, 1098–104 (2012). [PubMed: 22913530]
- (77). Zhang F & Finkelstein J Inconsistency in race and ethnic classification in pharmacogenetics studies and its potential clinical implications. *Pharmgenomics Pers Med* 12, 107–23 (2019). [PubMed: 31308725]
- (78). PharmVar. Allele Designation Criteria and Evidence Levels <https://www.pharmvar.org/criteria>. Accessed 06 April 2020.
- (79). PharmVar CYP2C19 Gene Expert Panel roster <https://www.pharmvar.org/expert-panels>. Accessed 06 April 2020.
- (80). NCBI Reference Sequences database <https://www.ncbi.nlm.nih.gov/refseq/>. Accessed 06 April 2020.

- (81). Locus Reference Genomic (LRG) project <https://www.lrg-sequence.org/>. Accessed 06 April 2020.
- (82). PharmVar Standards <https://www.pharmvar.org/genes>. Accessed 06 April 2020.
- (83). PharmVar CYP2C19 Change Log [https://a.storyblok.com/f/70677/x/01a6f89a17/change-log\\_cyp2c19\\_v1-9.pdf](https://a.storyblok.com/f/70677/x/01a6f89a17/change-log_cyp2c19_v1-9.pdf). Accessed 06 April 2020.
- (84). Drogemoller BI et al. Characterization of the genetic profile of CYP2C19 in two South African populations. *Pharmacogenomics* 11, 1095–103 (2010). [PubMed: 20712527]
- (85). Novkovic M, Matic D, Kusic-Tisma J, Antonijevic N, Radojkovic D & Rakicevic L Analysis of the CYP2C19 genotype associated with bleeding in Serbian STEMI patients who have undergone primary PCI and treatment with clopidogrel. *Eur J Clin Pharmacol* 74, 443–51 (2018). [PubMed: 29260275]
- (86). Gaedigk A et al. The Evolution of PharmVar. *Clin Pharmacol Ther* 105, 29–32 (2019). [PubMed: 30536702]
- (87). Numanagic I et al. Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. *Nat Commun* 9, 828 (2018). [PubMed: 29483503]
- (88). 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]
- (89). Astrolabe Tool for allele calling. Accessed 06 April 2020
- (90). 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]
- (91). Regan JF et al. A rapid molecular approach for chromosomal phasing. *PLoS One* 10, e0118270 (2015). [PubMed: 25739099]
- (92). Boone EC et al. Long-distance phasing of a tentative ‘enhancer’ SNP with CYP2D6 star allele definitions. *Frontiers Pharmacogenetics* (2020). doi.org/10.3389/fphar.2020.00486
93. () PharmVar Read Me for CYP2C19 [https://a.storyblok.com/f/70677/x/74372b2381/gene-info\\_cyp2c19\\_v2-2.pdf](https://a.storyblok.com/f/70677/x/74372b2381/gene-info_cyp2c19_v2-2.pdf). Accessed 06 April 2020.
- (94). PharmVar Structural Variation for CYP2C19 [https://a.storyblok.com/f/70677/x/9225fa2a03/variation\\_cyp2c19.pdf](https://a.storyblok.com/f/70677/x/9225fa2a03/variation_cyp2c19.pdf). Accessed 06 April 2020.
- (95). PharmVar P450 Nomenclature site - Archive [https://www.pharmvar.org/htdocs/archive/index\\_original.htm](https://www.pharmvar.org/htdocs/archive/index_original.htm). Accessed 06 April 2020.
- (96). CPIC Guidelines <https://cpicpgx.org/guidelines/>. Accessed 06 April 2020.
- (97). 10xGenomics Linked-reads Genomics <https://www.10xgenomics.com/linked-reads/>. Accessed 06 April 2020.
- (98). HGVS Sequence Variant Nomenclature. Accessed 06 April 2020



### Figure 1. Overview of the gene locus and allelic variation

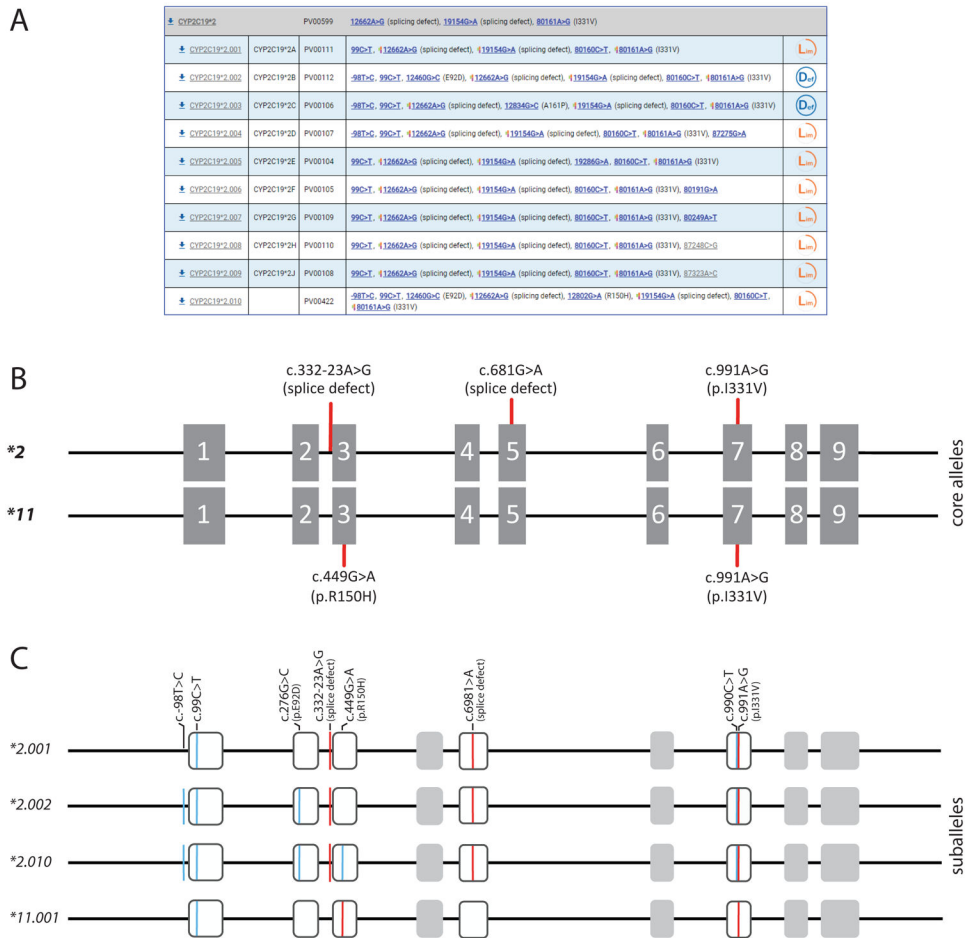
**Panel A** provides a graphical overview of the *CYP2C* gene locus containing *CYP2C18*, *CYP2C19*, *CYP2C9*, and *CYP2C8*. The latter is encoded on the reverse strand (arrow) while the other three genes are encoded on the forward strand. *CYP2C19* is composed of nine exons, covering 92.9 kb. Rare whole and partial gene deletion events that include at least exon 1 of the *CYP2C19* gene have been designated by PharmVar as *CYP2C19*\*36 and \*37, respectively. Whole gene deletions may include the neighboring *CYP2C* gene(s) and even extend beyond the *CYP2C* gene locus. Partial deletions affect different regions of *CYP2C19* and may affect *CYP2C18* and/or *2C9* as well. **Panel B** summarizes the core sequence variations defining *CYP2C19*\*2 through \*10 and \*17.



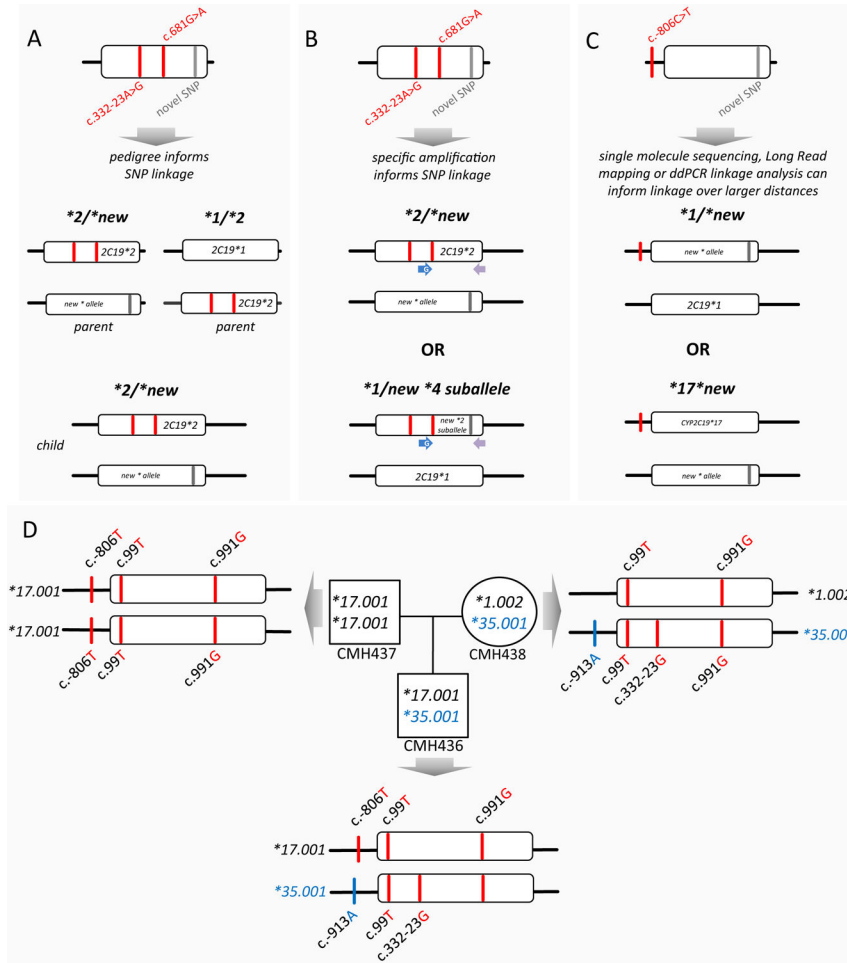
**Figure 2. Allele default assignment strategy used by many testing platforms**

Many pharmacogenetic test platforms only comprise the more commonly observed SNVs and do not include all known SNVs, or star (\*) alleles that are defined by PharmVar. Consequently, some alleles may not be identified or receive an assignment by ‘default’. It is therefore important to know which SNVs are tested in order to fully understand how alleles were called and translated into phenotype, as well as to appreciate a test’s limitations. **Panel A** visualizes the two known *CYP2C19*\*4 subvariants which are nonfunctional due to c.1A>G (p.M1V). The *CYP2C19*\*4.002 subvariant also carries c.-806C>T which is the core SNV defining the increased function \*17 allele. Thus, if testing only includes c.-806C>T and not c.1A>G, a *CYP2C19*\*4 allele will be called \*17 potentially leading to an incorrect phenotype assignment. **Panel B** illustrates that g.12662A>G (c.332-23A>G), which causes aberrant splicing, occurs on both, *CYP2C19*\*2 and \*35. If not tested, the latter will be defaulted to a *CYP2C19*\*1 assignment that may also lead to an incorrect phenotype assignment. **Panel C** shows the comparison of *CYP2C19*\*2, \*4, \*17 and \*35, created with the PharmVar CAVE tool. Blue boxes indicate the presence of a core SNV on all suballeles, while the gray box indicates that the core SNV is not present on all suballeles. The function (☐) symbol indicates that a core SNV alters function and the PharmVar (☐) symbol highlights that a core SNV is unique to a star allele. SNV positions refer to genomic coordinates on the NG\_008384.3 reference sequence (with the ATG start codon being +1).





**Figure 3. Overview of core alleles, suballeles, and the graphical Core Allele ViewEr (CAVE)**  
**Panel A** shows the *CYP2C19\*2* core allele definition (gray bar). Core SNVs, PharmVar ID (PVID), and evidence level is shown for each allele. All currently defined suballeles are displayed. Legacy allele designations are cross-referenced (e.g., *\*2.001* corresponds to *\*2A*). Note that g.12662A>G (c.332-23A>G) was added to all suballeles, when the gene was transferred to PharmVar (Table 3). **Panel B** is a graphical representation of the *CYP2C19\*2* and *\*11* core alleles and their core SNVs. c.991A>G (p.I331V) is present on both alleles, while c.332-23A>G and c.681G>A (splice defects) are the *CYP2C19\*2* core SNVs and c.449G>A (p.R150H) is the core SNV defining *CYP2C19\*11*. Core SNVs are shown by red lines; gray boxes represent the nine exons (scale is approximated). **Panel C** shows three of the ten *CYP2C19\*2* suballeles defined to date and the only *CYP2C19\*11* haplotype. Of note, the core SNV of the latter (c.449G>A (p.R150H)) is also present on the *CYP2C19\*2.010* suballele. Core SNVs (causing an amino acid change or aberrant splicing) are shown in red, all other SNVs are highlighted in blue (c.449G>A (p.R150H)) is not a *\*2* core SNP because it is not found on all suballeles.



**Figure 4. Experimental approaches for phasing SNVs to establish haplotype**  
**Panels A and B** depict a subject who is heterozygous for three SNVs of which two designate the *CYP2C19\*2* allele. Whether the novel SNV is in *cis* or *trans* with the *CYP2C19\*2* SNVs can be informed by e.g., inheritance (panel A), or experimentally determined using e.g., allele-specific (long-range) PCR, followed by sequencing (panel B), given that the distance among the novel SNV is not exceeding PCR amplification limitations. **Panel C** provides an example of a subject who is heterozygous for two SNPs, one of which designates the *CYP2C19\*17* allele. In this instance, the phase of the novel SNV may be inferred by inheritance (not shown) or experimentally determined by single molecule sequencing or utilizing ddPCR-based SNP-linkage analysis. **Panel D** details a submission for the *CYP2C19\*35* allele. The haplotype of the *CYP2C19\*35* suballele was determined from whole genome sequence data of each member of the trio. This data revealed an additional SNV on the *CYP2C19\*35* haplotype, as well as provided evidence to support changing the allele’s evidence level to ‘Def’.

**Table 1**Online *CYP2C19* Resources - Links to Sites and Online Resources Referenced Throughout the Review

Sources and References	References
<b>PharmVar</b>	
<i>CYP2C19</i> important gene information <ul style="list-style-type: none"> <li>Read Me Document</li> <li>Change Log Document</li> <li>Structural Variation Document</li> </ul>	(83, 93, 94)
Standards	(82)
Allele Designation and Evidence Level Document	(78)
<i>CYP2C19</i> Gene Expert Panel Roster	(79)
P450 Nomenclature Site – Archive	(95)
<b>PharmGKB</b>	
<i>CYP2C19</i> gene page	(43)
Gene-Specific Information Tables for <i>CYP2C19</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>	(44)
<ul style="list-style-type: none"> <li><i>CYP2C19</i> Drug Label Annotations</li> </ul>	(15)
<b>CPIC</b>	
Guidelines	(96)
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>	(18)
<b>Other resources</b>	
Drug Interactions Flockhart Table™	(30)
FDA Pharmacogenomic Biomarkers in Drug Labeling	(16)
FDA Table of Pharmacogenetic Associations	(17)
NCBI Reference Sequences Database	(80)
Locus Reference Genomic (LRG) Project	(81)
10X Genomics (Linked-Reads Genomics)	(97)
HGVS Nomenclature	(98)

**Table 2**

Novel allele(s) and confirmed suballele(s)

Core Allele Designation	Novel alleles/suballeles
*1	*1.004, *1.005
*2	*2.0010
*36	Complete deletion of <i>CYP2C19</i> gene
*37	Partial deletion of <i>CYP2C19</i> gene (i.e. deletions that include at least the entire exon 1)
Submissions for known alleles; original haplotype was confirmed, and evidence level raised to 'Def'	
<i>*1.002, *2.002, *15.001, *17.001, and *35.001</i>	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3.**

Summary of edits and changes during the transition into the PharmVar database and notable changes made thereafter

Reason	Change	Affected Alleles
Standardization	Intronic SNVs were removed	<i>*2C, *2D, *3B, *18, and *19</i>
	Retired star alleles removed	<i>*20, *21</i>
	Comment 'g.12662A>G is likely part of all *2 alleles' was removed	<i>*2A, *2C, *2B, *2E, *2F, *2G, *2H, and *2J</i> g.12662A>G was added to allele definition
	Comments removed	<i>*3B, *11, *16, and *30</i>
Other	Reassigned to <i>*1.006</i>	<i>*27</i>
	Correction	<i>*27</i> (99T>C was revised to 99C>T)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript