

Important Information

***CYP2D6* has been transitioned from the original “Nomenclature Website” (www.cypalleles.ki.se/) to PharmVar on Sept 26, 2017 (the last posted version of the *CYP2D6* page is available through the ‘Archive’ link). Current gene information can be accessed through the ‘genes’ tab on the ‘menu’ bar.**

A number of allelic variants, especially those defined early on, have not been fully sequenced and were defined based on exon and very limited up and downstream sequences. Also, some haplotypes may be inferred without experimental validation. These alleles may carry single nucleotide variants (SNVs) in introns and/or up- and downstream regions that have not been captured.

Gene Region Mapped

Allele definitions contain variants positioned between -1619 and 4482 of the NG_008376.3 RefSeq. There are known SNPs upstream and downstream of this region, however, these have not been associated with a change in function. Note that new submissions will need to include sequence covering a minimum of 1600 bp of upstream (including the -1584C>G SNP) and 250 bp of downstream region.

There is evidence that a distant enhancer SNP (rs5758550) located about 115 kb downstream of the *CYP2D6* gene locus impacts expression levels. Haplotype definitions, at this time, do not include the enhancer SNP.

Two homopolymer strings of A’s in the upstream region (2943-2964 (corresponding to -1258 to -1237) and 3098-3107 (corresponding to -1103 to -1094)) and the 133 bp within these strings are notoriously difficult to resolve by Sanger sequencing (and are almost always submitted with a gap for this region). For many haplotype definitions it is therefore unclear or unknown whether this region contains any sequence variations including the exact number of A’s within 2943-2964.

Coordinates

Coordinates in the PharmVar database are available in reference to a number of sequences, including Human Genome Assemblies GRCh37 and GRCh38, M33388, NG_008376.3 and NM_000106.5. The following table summarizes important information regarding differences among respective sequences.

In this ReadMe document coordinates are counted from the “A” in the ATG translation start codon as +1 for NG_008376.3 and M33388 in brackets unless indicated otherwise.

 **PharmVar feature:** No more counting. Get coordinates from the sequence start or the ATG start codon with one click! The default setting is NG_008376.3 counting the “A” in the ATG start codon as +1. The  buttons link out GenBank to access respective sequences.

Table 1

Sequence	notes
CYP2D6	Encoded on the negative strand
M33388	SNPs were annotated in reference to this sequence on the Nomenclature Website. M33388, however, contains sequencing errors (601insC, 1289-90GC>CG, 1328delG and 1440delC) shifting position numbers when compared to other sequences. For example, the G>A SNP defining the <i>CYP2D6</i> *4 allele is at position 1846 according to M33388, but 1847 on NG_008376.3.
NG_008376.3	The <i>CYP2D6</i> Reference Sequence (RefSeq) is identical to AY545216 and matches the <i>CYP2D6</i> *1 reference allele. This RefSeq is used in the PharmVar database. PharmVar submissions must be annotated using this RefSeq.
NG_008376.4	Released August, 2018. This RefSeq contains an additional 819bp of upstream 1540bp of downstream sequence compared to NG_008376.3, but is otherwise identical.
LRG_303	Locus Reference Genomic (LRG) sequence. LRGs never change once defined. LRG_303 matches the NG_008376.4 RefSeq.
NM_000106.5	transcript RefSeq
GRCh37	The <i>CYP2D6</i> sequence in GRCh37 matches the <i>CYP2D6</i> *2 allele, which has multiple SNPs compared to the <i>CYP2D6</i> *1 reference allele. When reporting * (star) alleles using GRCh37, these differences need to be accounted for. For example, rs16947 (M33388 2850C>T; NG_008376.3 2851C>T) is present on <i>CYP2D6</i> *2 and numerous other haplotypes, and is not identified as a variant when compared to GRCh37.
GRCh38	The <i>CYP2D6</i> sequence matches the NG_008376.3 RefSeq.

Sequence variants

Sequence variants detailed in the PharmVar database comprise single nucleotide polymorphisms (SNPs) and insertions and deletions of a single or multiple nucleotides (indels), all of which are referred to as single nucleotide variants (SNVs).

Some SNVs are unique to an allele and only occur on a single haplotype, e.g. 2550delA (2549delA) is unique to the *3 haplotype. Other SNVs occur on numerous alleles, e.g. 2851C>T (2850C>T) is found on *2, *8, *11, *17 and many more, and 1022C>T (1023C>T) is part of the *17, *40, *58 and *64 haplotypes.

Variation View

By clicking on a SNV, the variation view window will slide in. This view provides SNV positions across all sequences in both count modes and the link to the NCBI dbSNP identifier (rs number; note that some SNVs may not have been allocated a dbSNP identifier). There is also a **bar providing the option to display all haplotypes with the selected variant**.

Core allele definitions

For many star alleles, there are a growing number of so-called suballeles all sharing one or more 'key' defining sequence variant(s) (see ALLELE DESIGNATION AND EVIDENCE LEVEL CRITERIA document for details). Although suballele information is valuable, for example test design (sequence or SNP-panel based) and the interpretation of test results, the distinction of suballeles is not necessary for phenotype prediction because **all alleles under a star number are assumed to be functionally equivalent**. Thus, even if a test is capable of distinguishing suballeles, these are generally simply reported as e.g. *CYP2D6**2 or *4, etc.

PharmVar and the PharmGKB have collaboratively developed core allele definitions. **Only sequence variations which change an amino acid or impact on function by changing expression levels or interfere with splicing and are present in all suballeles within an allele group, are part of the core allele definition. With this rule-based system suballeles are collapsed into a single 'core' definition representing all suballeles categorized under a star (*) number.**

A core allele has its own unique PVID.

A core SNV is a sequence variation that is part of a core allele definition.

Core allele definitions are highlighted by a grey background in the *CYP2D6* gene page 'Table View'. Suballeles are labeled as e.g. *CYP2D6**2.001, *CYP2D6**2.002, etc. and are shown under an allele's core definition with alternating white and blue backgrounds and core SNVs are highlighted with the PharmVar logo for easy identification. Function is shown in the core allele bar representing all alleles listed under that core allele definition.

For example, there are numerous *CYP2D6**2 suballeles listed. Of all the SNVs found on these suballeles, only two fulfill the rules and are shared among all suballeles, namely 2851C>T (R296C) and 4182G>C (S486T). Therefore, these two SNVs constitute the *CYP2D6**2 core allele definition. The *CYP2D6**2 core allele definition is shown below along with the first six defined suballeles.

CYP2D6*2		PV00456	2851C>T (R296C), 4181G>C (S486T)	normal function →
CYP2D6*2.001	CYP2D6*2A	PV00129	-1584C>G, -1235A>G, -740C>T, -678G>A, 214G>C, 221C>A, 223C>G, 227T>C, 232G>C, 233A>C, 245A>G, 310G>T, 745C>G, 842T>G, 1662G>C, 2851C>T (R296C), 3385A>C, 3585G>A, 3791C>T, 4181G>C (S486T), 4482G>A	 deposited by Gaedigk et al and by Nofziger Johansson et al, 1993 Panaserat et al, 1994 Raimundo et al, 2000 Sakuyama et al, 2008
CYP2D6*2.002	CYP2D6*2B	PV00150	1038C>T, 1662G>C, 2851C>T (R296C), 4181G>C (S486T)	 Marez et al, 1997
CYP2D6*2.003	CYP2D6*2C	PV00149	1662G>C, 2471T>C, 2851C>T (R296C), 4181G>C (S486T)	 Marez et al, 1997
CYP2D6*2.004	CYP2D6*2D	PV00152	2851C>T (R296C), 4181G>C (S486T)	 Marez et al, 1997
CYP2D6*2.005	CYP2D6*2E	PV00151	996C>G, 1662G>C, 2851C>T (R296C), 4181G>C (S486T)	 Marez et al, 1997
CYP2D6*2.006	CYP2D6*2F	PV00147	1662G>C, 1725C>T, 2851C>T (R296C), 4181G>C (S486T)	 Marez et al, 1997

Important: since GRCh37 matches *CYP2D6**2, the 2851C>T and 4181G>A core SNPs are not displayed when GRCh37 is chosen. These SNPs are also not displayed for other core alleles.

To view *CYP2D6* core alleles, we strongly recommend to choose the RefSeq NG_008376.03 or GRCh38 as display setting.

For *CYP2D6**3, only 2550delA causing a frame shift is shared among the two known suballeles. Therefore, this SNV is the sole variant in the *3 core allele definition.

CYP2D6*3		PV00455	2550delA (259frameshift)	no function ×
CYP2D6*3.001	CYP2D6*3A	PV00221	2550delA (259frameshift)	 Kagimoto et al, 1990 Scantamburlo et al 2017
CYP2D6*3.002	CYP2D6*3B	PV00220	1750A>G (N166D), 2550delA (259frameshift)	 Marez et al, 1997

 CAVE - the PharmVar Comparative Allele ViewEr

PharmVar has developed the CAVE tool to easily compare core alleles, visualize which core SNVs are present in alleles of interest and to identify SNVs that are unique to selected alleles.

To access CAVE switch from Table View to Compare View



A sequence variation that is part of a core allele definition may be unique or occur in two or more core alleles. It may, however, also be found in some other suballeles. In contrast, some non-core SNVs may be unique and can tentatively be utilized to identify an allele of interest or discriminate it from others. As mentioned above, the CAVE tool uses core allele definitions, and hence unique SNVs that are not part of any core definition are not displayed.

The 100C>T SNV is one prime example to illustrate the scenario of a SNVs being part of multiple core allele definitions including *CYP2D6**10, *36, *49, *100 and *114 to name a few. As detailed below in example 1, it is, however, also present on all but one *CYP2D6**4 suballele and found on the vast majority of *CYP2D6**4 haplotypes observed in global populations.

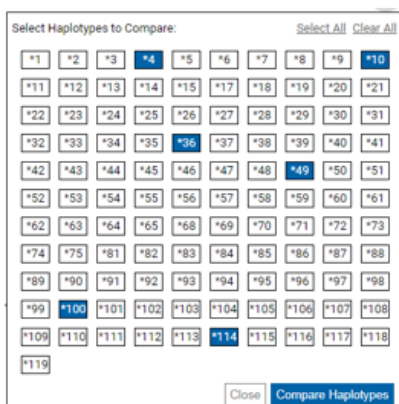
 **CAVE comparisons can be downoladed into an Excel spreadsheet using the 'Download Comparison table' button.**

The following examples demonstrate of how to use CAVE and how information is displayed.

Example 1:

Select core alleles for comparison in the **selection pad**. For this example, we have selected five core alleles all having 100C>T in their core allele definition, i.e. *10, *36, *49, *100 and *114. *CYP2D6**4 was selected because 100C>T is present on many of its suballeles, but is not a core SNV.

Allele Selection Pad



 Click compare haplotypes

CAVE graphically displays the core SNVs that are found in the selected alleles

- 100C>T is shown in **blue** for *10, *36, *49, *100 and *114 indicating that it is part of their respective core allele definitions. This SNV is shown in **gray** for *4 indicating that this SNV is not part of this allele's core allele definition, because it is not found on all *4 suballeles. Since 100C>T is not unique to *10, this allele cannot unequivocally be identified by genotyping for 100C>T; the presence of other variants carrying 100C>T must be ruled out by testing for their respective core SNVs.
- Similarly, 4181G>C is part of the *10, *36, *49, *100 and *114 core allele definitions, but not *4 and thus is shown in **blue** and **gray**, respectively. This SNV does not identify any particular allele, because it is part of many haplotypes. Also, according to current knowledge, this SNP does not impact function although it causes an amino acid change.

	*4	*10	*36	*49	*100	*114
100C>T						
1612T>A						
1759G>A						
1847G>A						
2829delC						
2851C>T						
Ex9Conv ⓘ						
4181G>C						

Variant is present
 Variant is present on some suballeles
 Variant is unique for selected haplotypes
 Variant alters function

Example 2:

Select core alleles for comparison in the **selection pad**. For this example, we have selected 8 core alleles all having 2851C>T as a core SNV, i.e. *2, *11, *12, *14, *31, *35, *41 and *69.

- 2851G>A is shown in **blue** for all alleles indicating that these are part of their respective core allele definitions. Since this SNV is not unique to any of the selected alleles, these cannot be discriminated by genotyping for these SNVs.
- 4181G>C is also shown in **blue** because this SNP is part of the core definition of the selected alleles. *CYP2D6**11, *12, *31 and *35 each have a SNP that is unique for to these alleles, i.e. these SNVs can be utilized to unequivocally identify these alleles.

	*2	*11	*12	*14	*31	*35	*41	*69
31G>A						Variant is present		
100C>T								Variant is present
124G>A			Variant is unique for selected haplotypes					
882G>C		Variant is unique for selected haplotypes						
1759G>A				Variant is unique for selected haplotypes				
2851C>T	Variant is present	Variant is present	Variant is present	Variant is present	Variant is present	Variant is present	Variant is present	Variant is present
2989G>A							Variant is present	Variant is present
4043G>A					Variant is unique for selected haplotypes			
4181G>C	Variant is present	Variant is present	Variant is present	Variant is present	Variant is present	Variant is present	Variant is present	Variant is present

- Variant is present
- Variant is present on some suballeles
- Variant is unique for selected haplotypes
- Variant alters function

Function

The function of an allele is shown according to the information curated by the PharmGKB and Clinical Pharmacogenetic Implementation Consortium (CPIC) available at <https://www.pharmgkb.org/page/cyp2d6RefMaterials>. Allele functionality tables are created as part of a more formalized CPIC guideline process and are periodically updated. Additional information may have become available since the functionality table has been created. For some alleles, the assigned activity may differ from that shown on the archived Nomenclature pages. Extrapolation of the activity of a particular allele to particular substrates should be done with caution due to potential unknown substrate-specific activities of the allele.

CYP2D6 Structural Variation document

This document provides detailed information regarding structural variants including gene deletions and duplication (copy number variation, CNV), conversions and structural rearrangements between CYP2D6 and its pseudogenes.

Important: Gene duplications, hybrid genes and structural variants are not shown in the PharmVar database as separate entries, but are described in detail in the CYP2D6 Structural Variation document. Hybrid genes are listed providing the general structure and referral to the CYP2D6 Structural Variation document.

References

The references provided in the PharmVar database and the ReadMe document include the citation in which an allele was first published. For some alleles additional reference(s) describe important updates and/or information regarding function. The reference list is not intended to provide a complete bibliography for an allele. Haplotypes not published elsewhere are listed as “deposited by”.

Allele frequencies

CYP2D6 allele frequency tables have been developed for CPIC guidelines and are available through the PharmGKB at <https://www.pharmgkb.org/page/cyp2d6RefMaterials>. A comprehensive list of frequencies and references can be found in the *CYP2D6* allele frequency Table in the ‘source’ tab. These tables are periodically updated.

Changes and Edits

A number of changes and edits have been made to the annotations on the original P450 Nomenclature site to standardize annotations across genes and correct errors. Please see the [Change Log](#) document for details.