Important Information

***CYP3A4* has been transitioned from the original "Nomenclature Website" (www.cypalleles.ki.se/) to PharmVar on Sept 26, 2017 (the last posted version of the *CYP3A4* page is available through the 'Archive' link). Current gene information can be accessed through the 'gene' tab on the 'menu' bar.**

Several allelic variants may not have been sequenced for intronic flanking regions or may have very limited information for up and downstream sequences, or relevant intronic sequence variants such as rs35599367 defining *CYP3A4*22*. Also, for some alleles, the defined haplotype is inferred and was not experimentally confirmed. Such alleles may carry single nucleotide variants (SNVs) in exons, flanking intronic and/or up- and downstream regions that have not been captured when the allele was first defined.

Gene Region Mapped

Allele definitions contain variants positioned between 500 bp of upstream and 250 bp of downstream region of NG_008421.1 (counting from the start and stop codons, respectively). Note that new submissions must include respective upstream and downstream regions. Intronic variants are not shown as part of haplotype definitions unless the variant impacts function.

Coordinates

Coordinates in the PharmVar database are available in reference to several sequences including Human Genome Assemblies GRCh37 and GRCh38, the genomic NG_008421.1 RefSeq and the NM_017460.6 transcript reference. **Table 1** 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 NM_017460.6 unless otherwise stated.

PharmVar Feature: Get coordinates from the sequence start or the ATG start codon with one click! The default setting is NG_008421.1 counting the "A" in the ATG start codon as +1. The ⓘ buttons link out to GenBank to access respective sequences.

Table 1

sequence	notes
CYP3A4	Encoded on the negative strand
NG_008421.1	This *CYP3A4* Reference Sequence (RefSeq) matches the *CYP3A4*1* allele definition (see below regarding SNV at c.-392). **This sequence is used as reference to define allelic variation (star alleles)** **PharmVar submissions must be annotated using the NG_008421.1 sequence**
n/a	Locus Reference Genomic (LRG) sequence. LRGs never change once defined **There is no LRG for *CYP3A4***
NM_017460.6	Transcript RefSeq; MANE Select v0.92 The MANE Select represents the gene's transcript that is representative of biology, well-supported, expressed and is highly conserved. This transcript matches GRCh38 and is 100% identical with its RefSeq for 5'UTR, CDS, splicing and 3'UTR
GRCh37	The *CYP3A4* sequence matches the *CYP3A4*1* allele definition (NG_008421.1)
GRCh38	The *CYP3A4* sequence matches the *CYP3A4*1* allele definition (NG_008421.1)
AF280107.01	Legacy reference sequence; this sequence differs from NG_008421.1. Most notably, the SNV at position -392 was annotated as A>G in the past using AF280107.01 (see archive), while the current RefSeq NG_008421.1 shows this SNV as c.-392G>A (see more details below)

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) to which PharmVar collectively refers to as single nucleotide variants, or SNVs.

Many SNVs are unique to an allele and only occur on a single haplotype, e.g., c.1461insA (rs67666821) is only found on *CYP3A4*20*. There are only two SNVs which have been described to be part of two or more haplotypes. c.-392G>A (rs2740574) is part of numerous haplotypes, and c.1026+12G>A (rs2242480) is an intronic SNV with emerging evidence that it may cause altered function.

c.-392G>A (rs2740574): The SNV at c.-392 was defined as A>G in the past (using the now discontinued reference sequence AF280107.01). Since this SNV is defined as G>A on the current RefSeq (NG_008421.1), all allele definitions were updated accordingly. Specifically, all alleles with the c.-392A>G SNV now match the RefSeq and are thus no longer showing the variant, while all other alleles gained c.-392G>A. Note that the 'legacy nomenclature' showed *CYP3A4*1A* without a SNV and *CYP3A4*1B* having c.-392A>G. Now, *CYP3A4*1B* corresponds to *CYP3A4*1.001* (having a G at c.-392) and *CYP3A4*1A* allele corresponds to *CYP3A4*1.002* (having an A at c.-392). Legacy names are cross-referenced with their new PharmVar names on the gene page.

***CYP3A4*1G*:** The defining variant of this allele is **c.1026+12G>A (rs2242480)** in intron 10 (this variant is also part of several other star alleles). The *CYP3A4*1G* allele has been associated with increased CYP3A4 activity, while others reported associations with decreased activity. It is unclear though from published evidence if c.1026+12G>A is indeed the underlying cause. Observed altered activity might be caused by another SNV(s) that is in linkage with c.1026+12G>A or be confounded by activity encoded by the other genes in the *CYP3A* locus. Although all other intronic SNVs of unknown function have been removed from allele definitions, **PharmVar continues to display *CYP3A4*1G* under its 'legacy' name, as well as displays c.1026+12G>A in other haplotypes, until additional evidence is available**. If new information supports c.1026+12G>A being a function altering SNV, it will receive core SNV status, and consequently, **1G* will receive its own star allele designation. Otherwise, *CYP3A4*1G* will be retired and c.1026+12G>A removed from all allele definitions.

Missing information for c.522-191C>T (rs35599367, core SNV of *CYP3A4*22*) and/or c.1026+12G>A (rs2242480, defining *CYP3A4*1G*): For many alleles there is no published information for these intronic SNVs; such alleles received an evidence level of LIM. An allele received an evidence level of MOD, if linkage data obtained from the 1000 (1K) Genomes Project data strongly supported the absence or presence of these intronic SNV(s). For example, *CYP3A4*4* received an evidence level of MOD because the two intronic SNVs were not found on any sample in the 1K Genomes Project that was positive for the **4* core SNV. Alleles not represented in the 1K Genomes Project received an evidence level of LIM.

All future submissions must provide information for c.522-191C>T and c.1026+12G>A. PharmVar also encourages submissions for existing allele definitions to confirm the status of these SNVs. "

Variation View

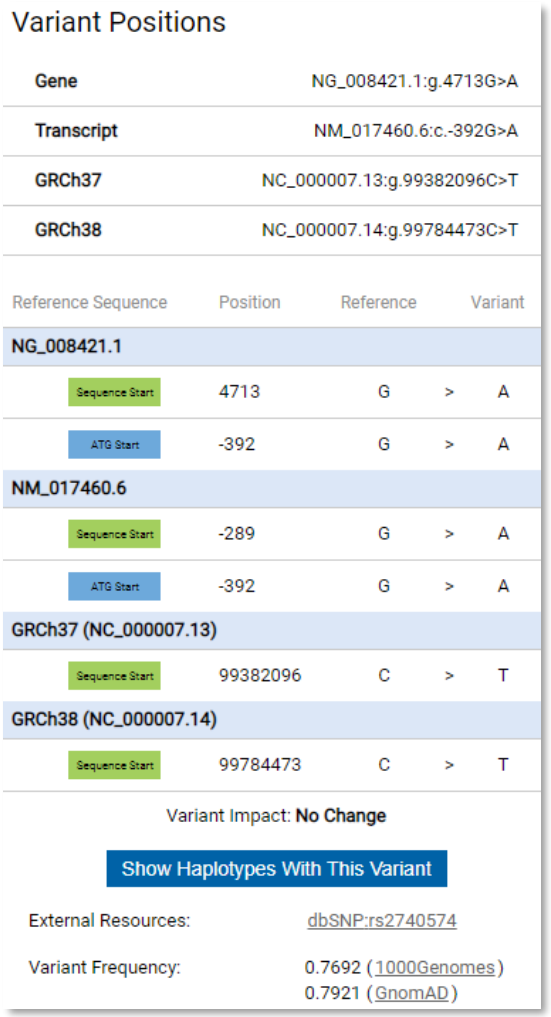
By clicking on a SNV, the variation view window will slide in (the example shows the *CYP3A4* variant c.-392G>A that is present on numerous star alleles). This view provides SNV positions across all sequences, the link to the NCBI dbSNP identifier (rs number; note that some SNVs may not have been allocated a dbSNP identifier) as well as SNV frequency. There is also a **bar providing the option to display all haplotypes with the selected variant**.

The **top portion** of the variation window **displays SNV coordinates according to Human Gene Variation Society (HGVS)** nomenclature on the gene, transcript and genome (GRCh37 and GRCh38) levels. Coordinates are displayed as obtained through the NCBI Variation Services and only manually curated if no results are returned. Additional details on HGVS nomenclature can be found at https://varnomen.hgvs.org/recommendations/DNA/.

The **middle portion** of the variation window **displays SNV positions 'PharmVar-style'** on the gene, transcript, and genome (GRCh37 and GRCh38) levels giving positions for both count modes and detailing the reference and variant nucleotides.

Note that PharmVar consistently uses the 3' Rule (left alignment) to determine the positions of inserted or deleted nucleotide(s). HGVS and 'PharmVar-style' positions may differ for indels in some instances (such as the example shown) which is most likely explained by differences in sequence alignments. Also, PharmVar displays single nucleotide insertions as 'ins' while HGVS displays them as duplications or 'dup'. Additional details and examples are provided under 'Standards'.

The **bottom portion** of the variation window provides a filter option to display all variants with the selected SNV, the link to dbSNP (if rs number exists for SNV), as well as SNV frequency information.



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 Criteria and Evidence Level" for details). While suballele information is valuable for 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., *CYP3A4*1*. Currently, *CYP3A4*1* is the only allele with suballeles.

PharmVar has developed core allele definitions in collaboration with the PharmGKB. Only sequence variations which change an amino acid or impact 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**. Core alleles have their own unique PVID.

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

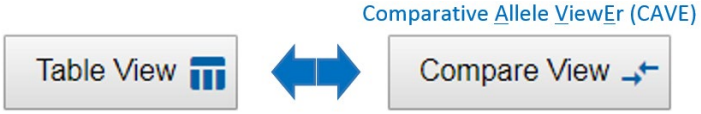
CYP3A4*1		PV01444			CPIC Clinical Function →
CYP3A4*1.001	CYP3A4*1B	PV01482		Def	Rebbeck et al. 1998
CYP3A4*1.002	CYP3A4*1A	PV01503	-392G>A	Def	Gonzalez et al. 1988 deposited by Gaedigk et al
CYP3A4*1.003	CYP3A4*1E	PV01483	-392G>A, -369T>A	Lim	Hamzeiy et al. 2002
CYP3A4*1.004	CYP3A4*1M	PV01492	-392G>A, -156C>A	Lim	Fukushima-Uesaka et al. 2004
CYP3A4*1.005	CYP3A4*1T	PV01477	-392G>A, 26022T>C	Lim	Fukushima-Uesaka et al. 2004
CYP3A4*1.006		PV01512	-392G>A, -301T>C	Def	deposited by Gaedigk et al
CYP3A4*1G	CYP3A4*1G	PV01486	-392G>A, 20239G>A	Def	Fukushima-Uesaka et al. 2004

Core allele definitions are highlighted by a gray background in the *CYP3A4* gene page 'Table View'. Suballeles are labeled as e.g., *CYP3A4*1.001, etc.* and are shown under an allele's core definition with alternating white and blue backgrounds.

The PharmVar Comparative Allele ViewEr (CAVE)

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.

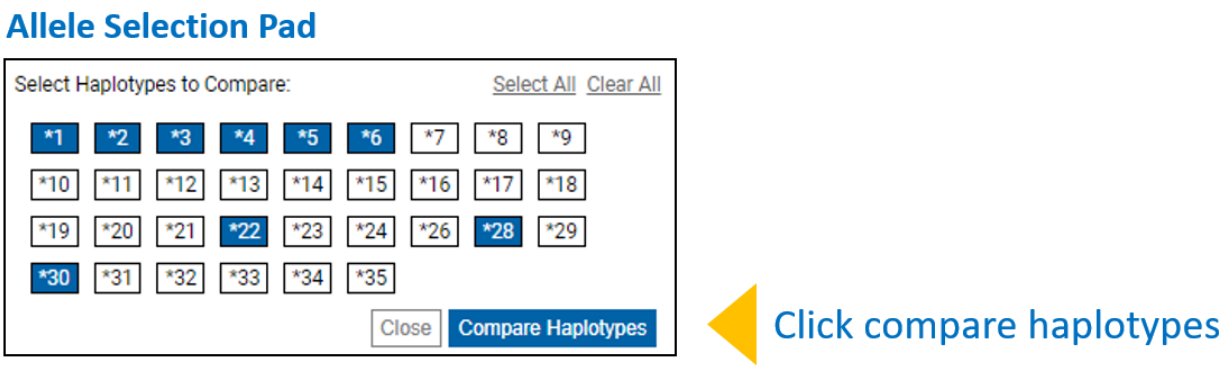
To access CAVE switch from Table View to Compare View



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 may 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.

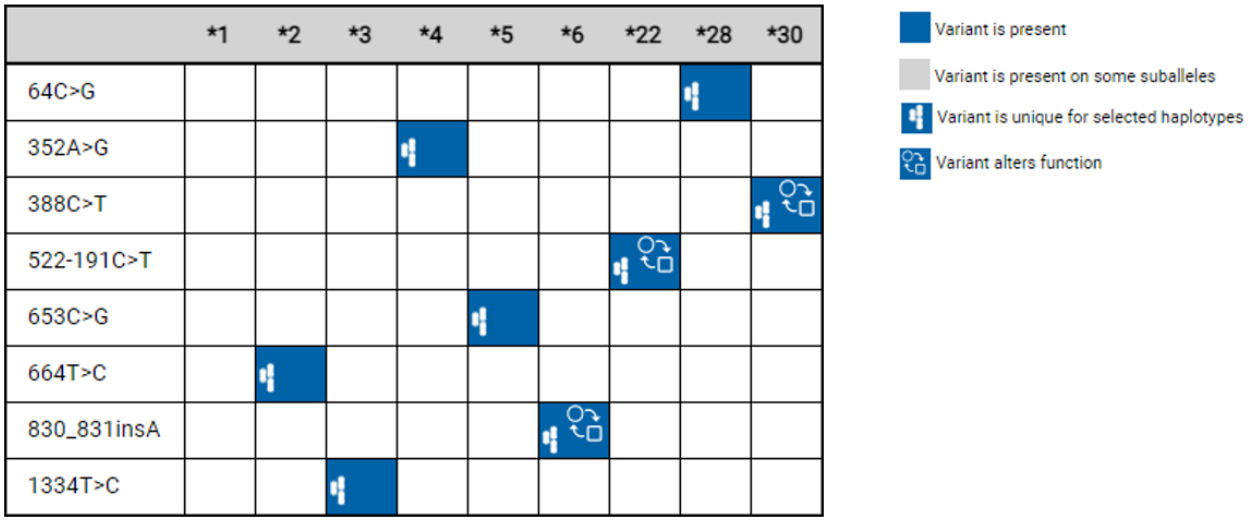
Examples:

Select core alleles for comparison in the **selection pad**. To illustrate CAVE for *CYP3A4*, the following alleles have been selected, i.e., **1, *2, *3, *4, *5, *6, *22, *28 and *30*.



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

- Each allele is defined by a unique core SNV as indicated in. SNVs unique to an allele are denoted by the PharmVar logo. Missense SNVs predicted to alter function are marked by the function symbol. c.522-191C>T of *CYP3A4*22* also has the function symbol as this allele has been consistently described in the literature to convey decreased function. Other SNVs have not been systemtically reviewed and therefore, no function annotations are provided; see more below regrading function.
- For this gene, there are no examples to date for a core SNVs being present on more than one star allele.



Variant Groupings

This feature allows PharmVar to display amino acid changes that are caused by two SNVs or a series of SNVs within a defined region that cause multiple amino acid changes.

There are no examples for this feature for *CYP3A4.*

Function

Function of *CYP3A4* alleles has not been curated by CPIC; therefore, alleles are shown as CPIC clinical function 'not available'. Alleles containing missense variants such as a premature stop codon or a frameshift are annotated as no function, however.

Allele Frequencies

CYP3A4 allele frequencies have not been systematically collected.

Variant Frequencies

PharmVar displays the frequency of a sequence variation in the **Variation Window** next to its link to dbSNP. The frequency of the SNV is provided for the gnomAD database (https://gnomad.broadinstitute.org/) and the 1000 Genomes Project (https://www.internationalgenome.org/). In some cases, information may only be available for one or the other or neither. These frequencies represent the global frequency of the SNV in each of these databases.

For example, the **Variation Window** for c.522-191C>T (rs35599367) defining *CYP3A4*22*, indicates that the frequency of this SNV is 0.03089 (3.089%) in the gnomAD and 0.0150 (1.50%) in the 1000 Genomes Project.

The frequency of a SNV does not necessarily reflect the frequency of a haplotype if it is part of two or more haplotypes. For example, c.1026+12G>A (rs2740574) is not unique to *CYP3A4*1G* and thus, the frequency of this SNV represents the sum of the frequency of all alleles with this SNV.

References

The references provided in the PharmVar database include the citation in which an allele was first published. For some alleles additional reference(s) describe important updates supporting the haplotype. The reference list is not intended to provide a complete bibliography for an allele. Haplotypes not published elsewhere are listed as "deposited by".

Changes and Edits

Several changes and edits have been made to the annotations on the original P450 Nomenclature site to standardize annotations across genes and correct errors. Any future changes will be logged on the **CHANGE Log** document.