

The Human Cytochrome P450 (CYP) Allele Nomenclature Database, which served as an authority for CYP allele designation since 2002, has transitioned to the **Pharmacogene Consortium (PharmVar)**. The major focus of PharmVar is to continue the mission by serving as a repository for allelic variation, providing an official and unified allele designation system for the Pharmacogenetics (PGx) community and facilitating the translation of genotype into phenotype and clinical implementation of PGx.

PharmVar will designate human pharmacogene variation and house allelic variants in the PharmVar database.

1. All submissions to PharmVar must use the submission form available on www.PharmVar.org and be submitted to submissions@PharmVar.org. Only complete submission requests will be accepted and processed.
2. PharmVar Gene Expert Panels will review each submission and make a recommendation to the PharmVar Steering Committee.
3. In this ‘Allele Designation Criteria and Evidence level’ document, we collectively refer to deviations from the RefSeq as “sequence variations”, including single nucleotide polymorphisms (SNPs) and nucleotide insertions and deletions (indels). Copy number variants (i.e. entire or partial gene deletions and duplications) hybrid genes (e.g. *CYP2D6/2D7* hybrids) and duplications containing non-identical gene copies (e.g. *CYP2D6*36+*10*) are being referred to as CNVs or structural variants.
4. The PGx nomenclature system **groups haplotypes and subvariants (suballeles) based on the presence of amino acid changes and/or functional non-coding** sequence variation and/or CNVs and structural variants **using star (*) designations** to facilitate phenotype prediction and clinical implementation.
5. Each haplotype (allele) designation will also receive a qualifier reflecting the **level of evidence** on which its designation is based (see below for details).
6. The new PGx nomenclature system assigns each haplotype a unique label as shown below that is comprised of the name of the gene (e.g. *CYP2C9*, *NUDT15*) followed by the major star (*) allele assignment and subvariant assignments (shown as a numerical string).

Major allele
CYP2C9*2.001
subvariant

Examples: **CYP2C9*3.001** (CYP2C9*3 subvariant 1)
CYP2D6*2.055 (CYP2D6*2 subvariant 55)

The most frequent, or one of the most frequent, versions of a haplotype is/has been often discovered first and therefore becomes the “prototype” defining all alleles grouped under that star (*) designation. There are a few cases, however, where the initially described haplotype may not be the most common. For example, emerging data now suggest that *CYP2D6*6A* (**6.001*) was designated first although data now suggest that it is less common compared to *CYP2D6*6B* (**6.002*). Also, subvariants discovered later may have additional or fewer SNPs than the variant allele initially discovered.

The new numerical strings were assigned in sequential order (and not re-ordered by frequency).

Examples: *CYP2D6*6A* (**6.001*), **6B* (**6.002*), **6C* (**6.003*) and **6D* (**6.004*)

*CYP2D6*10A* (**10.001*), **10B* (**10.002*) and **10D* (**10.003*). Since **10C* no longer exists **10D* became **10.003* using sequential numeric string assignments.

If two or more novel haplotypes and subvariants thereof are submitted to PharmVar simultaneously the haplotype with the least number of sequence variants will be designated *new.001, unless there is data justifying that a particular haplotype should be named as .001 (e.g. frequency).

7. Each haplotype receives a PharmVar-(PV) ID. The PV-ID is a unique identifier allowing the PharmVar database to track each haplotype; the PV-ID will be assigned by the PharmVar db team (see STANDARDS document for additional information).
8. The names for the corresponding proteins have a period between the name of the gene product and the number specifying the major haplotype (e.g. CYP2C9.3 and CYP2D6.2, etc.). An allele may not produce a full-length protein, for example, when frame shifts cause premature termination due to indels or aberrant splicing. In some instances, no protein will be expressed; for example, for *CYP2D6*5* (gene deletion), there will be no CYP2D6.5 expression product.
9. All haplotypes listed under a star designation are, to the best of current knowledge, of similar function. Thus, to be recognized as a major haplotype or allele, i.e. for an allele to receive a new star designation, **the new haplotype must contain at least one nucleotide change that results in an amino acid change, or has been experimentally shown to affect transcription, splicing, translation, activity, or lead to posttranscriptional or posttranslational modifications.**
 - a. If a known sequence variation that is associated with no function (e.g. 1846G>A in *CYP2D6*4* which leads to a splicing defect) is shown to exist together with any other sequence variation(s) (that do not recover activity), these will be listed as a subvariant (suballele) of the known haplotype causing loss of function.

Examples: **CYP2D6*4.001, CYP2D6*4.002**, etc.
 CYP2C9*3.001, CYP2C9*3.002, etc.
 - b. If a haplotype/allele harbors two or more sequence variations that are each known to obliterate function on their own, it will be listed under the major haplotype with the lowest star number. For example, a hypothetical allele with 1846G>A (defining *CYP2D6*4*) and 1707delT (defining *CYP2D6*6*) will be listed as a *CYP2D6*4* subvariant.
 - c. If a known sequence variation that is associated with decreased function (e.g. 100C>T defining *CYP2D6*10*) is shown to exist together with another non-synonymous sequence variation of known or unknown function, a new star number (major haplotype) will be assigned. For example, for an allele having both 100C>T (defining *CYP2D6*10*) and a new amino acid change known to decrease function on its own, the allele will receive a new star (*) designation. Likewise, an allele having 2615_2617delAAG (defining *CYP2D6*9*) that carries 4180G>C, which is not known to cause a change in function would also receive a new star (*) designation.

Example: **CYP2D6*new.001**, etc.
10. Allelic **haplotypes/alleles not qualifying as a major novel haplotype** as described in section 9 will be catalogued as subvariants. More specifically,
 - a. If a major haplotype carries additional sequence variation(s) that do not change an amino acid, do not interfere with splicing, nor convey any impact that may result in a change of function, these will be listed as subvariants. Subvariant status will also be assigned if the impact of these additional sequence variation(s) are unknown. A subvariant may be re-classified if future work demonstrates that sequence variation(s) found on a subvariant have a significant impact on function.
 - b. A major haplotype may have been defined by the presence of three variants, one being synonymous and two that cause amino acid changes. An allele lacking the synonymous SNP, but having the amino acid changing sequence variations, will be listed as a subvariant.
11. Novel sequence variation(s) identified by partial/incomplete gene sequencing (i.e. some but not all exons and/or required flanking sequences are sequenced) will not be accepted by PharmVar. This includes cases with functional sequence variations found within the characterized portions of the gene.

12. If a novel **haplotype was discovered by NGS**, details of the sequencing platform (e.g. Illumina, IonTorrent, PacBio etc.) and the type of sequencing performed (e.g. whole genome sequence, whole exome sequence, or other targeted panel along with other pertinent details i.e. kits and analysis software used) must be provided. NGS data must be of high quality in order for the submission to be acceptable. Variant calls must be made using sequence reads with base quality ≥ 10 and with an alignment quality ≥ 20 . Average coverage of the gene locus (that is submitted to PharmVar) must be at least 30X with a minimum coverage of 20X.

If a novel **haplotype was discovered by Sanger sequencing**, findings must be confirmed by repeat sequencing of a separate amplicon or a genotyping assay detecting the new sequence variation(s).

13. Structural variation and gene copy number

- a. For *CYP2D6*, *CYP2A6* and other genes under 10 kb in length, extra gene copies (xN) on the same chromosome (e.g. duplications or tandem arrangements) or other structural variants (e.g. hybrid genes or gene deletions), the entire structural variation must be fully characterized, i.e. both gene copies must be analyzed according to the requirements used for single gene copies per submission SOP. The methods of how the duplication or other structural variant was discovered and characterized must also be disclosed (e.g. XL-PCR, TaqMan CNV assay, etc.). Large deletions or duplications (affecting multiple genes including the pharmacogene of interest) may be defined as described in b.
- b. For genes over 10 kb in length, structural variation is often technically more difficult to characterize. For such cases, deletions or duplications must either be confirmed by two methods, e.g. Array Comparative Genomic Hybridization (aCGH) or other high-resolution array platform, TaqMan copy number assay(s), multiplex ligation-dependent probe amplification (MLPA) or other method(s) or demonstrated in at least two subjects. Exonic sequencing of novel duplicated alleles is strongly encouraged when feasible; however, it is not required for the submissions of CNVs or structural variants.

14. All submissions are kept confidential. Information will only be shared with the Gene Panel Experts and members of the PharmVar Steering Committee reviewing the submission. **Upon submission, new haplotype/allele designations will be kept confidential until date of publication or up to 6 months after a designation has been issued** (whichever comes first). Authors are requested to provide ePub, doi, PMID or a link to a preprint server of the manuscript/publication(s) as it becomes available. Authors are encouraged to submit their novel haplotypes/alleles to PharmVar prior to submission to a journal, especially because many editors request naming of new alleles. Journal publication is, however, not a requirement for submission and publication of alleles in the PharmVar database.

15. PharmVar **encourages the PGx community to submit additional information** for already designated variants with regards to function (*in-vitro* or *in-vivo* activity analyses) or frequency information at any time by email to submission@PharmVar.org. Please include the name of the designated variant as well as “update” in the subject line.

16. PharmVar also encourages submitters to obtain rs IDs for all SNVs that do not have an rs ID. It is, however, not a prerequisite for acceptance that all SNVs of a haplotype have rs IDs.

17. Finally, **usage of star (*) allele designations resembling those officially designated by the PharmVar Consortium is highly discouraged** due to the likely risk of confusion.

Levels of Evidence

For some genes it may be difficult to unequivocally determine linkage disequilibrium of variants (i.e. haplotype structure) across the entire gene. For smaller genes (under 10-15 kb), haplotype definitions are increasingly easy to experimentally verify. For example, a long-range (XL) PCR product can be subjected to single molecule NGS-based sequencing or, alternatively, an allele-specific XL-PCR can be Sanger sequenced to unequivocally determine an allele's haplotype. Inheritance patterns may also unequivocally demonstrate variant linkage disequilibrium in trios by demonstrating which sequence variations are passed from parent to child. For larger genes, however, haplotype definitions may be based on computational methods including phasing algorithms and haplotype linkage analyses. In such cases, pedigree information demonstrating variant linkage disequilibrium can be invaluable to infer haplotypes or validate computationally predicted haplotypes; pedigrees are, however, not always available or informative. In the future, novel technologies will likely become more widely available (e.g. the 10X Genomics platform which phases SNPs over long distances), enabling long-distance phasing and facilitating haplotype definitions.

Furthermore, NGS-based sequencing is increasingly utilized for the detection of sequence variation in research and for clinical applications. Although Sanger sequencing is still considered to be the Gold Standard and often required to validate NGS-based findings, PharmVar will accept NGS-based submissions, given that certain criteria are fulfilled.

Some allele/haplotype definitions are experimentally confirmed or demonstrated by inheritance while others are computationally inferred or are based on limited information. This may be important information for investigators using allele definitions as well as clinical implementation. It is emphasized, though, that all haplotypes accepted by PharmVar are based on high-quality data submissions; thus, the evidence level is not a measure of the quality of the data, but rather reflects the amount and nature of the data a haplotype definition is based upon.

Detailed information regarding the methods/approaches used for the characterization of a new haplotype, including which gene regions were interrogated and whether a haplotype was experimentally confirmed or computationally inferred, is being captured when an allele/haplotype is submitted to PharmVar. PharmVar accepts '**encore**' submissions for the same haplotype by investigators other than the original submitter or additional information from the original submitter to build a stronger body of evidence in support of a haplotype.

The symbols displayed on PharmVar indicate **DEFINITIVE, MODERATE and LIMITED** levels of support for a haplotype and reflect all submissions PharmVar has received for a haplotype (this 3-point system is modified from the classification system employed by ClinVar). This type of information was not systematically captured prior to PharmVar. For existing haplotype definitions, a literature review will be conducted for each gene to extract pertinent information to assign evidence levels for that gene's variant alleles. It will take time, however, to compile this information for already existing haplotypes (blank information on the gene page indicates that no evidence levels have been assigned yet).

Evidence levels will be assigned to alleles/haplotypes by respective gene expert panels as new submissions are being reviewed (panels may request additional information from submitters at their discretion). The evidence level descriptions below may not cover all possible scenarios and will be revised in the future as needed.

PharmVar encourages the PGx community to submit additional information for alleles/haplotypes with **LIMITED** and **MODERATE** evidence levels to complement information and raise their evidence levels to **DEFINITIVE**.

Evidence level **DEFINITIVE**

Gene fully characterized by Sanger sequencing or NGS-based sequencing.

One of the following criteria **MUST** be fulfilled:

- The subject's genotype unequivocally informs the new haplotype as follows:
 - a. The subject is homozygous for all SNVs
 - b. The subject is homozygous for all but one SNV
 - c. The subject is heterozygous for a single SNV
- Allele/haplotype is inferred by inheritance as follows:
 - a. Information from one or more pedigrees unequivocally demonstrates that the sequence variation(s) in the new allele/haplotype are inherited together. A cartoon providing information for informative family members must be provided (see example below)

It is recommended that novel SNVs discovered by WGA or WES are confirmed by Sanger sequencing or other method to confirm the presence and nature of the SNV.

- Allele/haplotype was experimentally confirmed as follows:
 - a. Sequencing of an allele-specific XL-PCR product that covers the entire or a set of allele-specific XL-PCR products;
 - b. Genotyping of all relevant SNPs using an allele-specific XL-PCR or a set of allele-specific XL-PCR products as template;
 - c. Using an amplification-refractory mutation system (ARMS) or similar approach to unequivocally demonstrate SNP linkage;
 - d. Linkage of sequence variation(s) in coding sequences can be demonstrated by sequence analysis of a cDNA clone, or an allele-specific template generated from cDNA;
 - e. Single-molecule NGS-based analysis (e.g. PacBio SMRT sequencing).
 - f. other appropriate means

Evidence level **MODERATE**

Gene fully characterized by Sanger or NGS sequencing

- Allele/haplotype is inferred by computational means as follows:
 - a. The population data used for phasing must fulfil the quality criteria described in 12 and must be publicly available. The new inferred haplotype must be observed at an allele frequency of at least 2% within the population data set when phasing against 100-500 subjects or 1% when phasing against ≥ 500 subjects (excluding the subject in which the novel allele/haplotype was first discovered). Subject ID(s) predicted to carry the new haplotype must be provided (e.g. the ID(s) of the Coriell sample(s) inferred to carry the novel allele/haplotype).

Note: Data from the 1000 Genomes Project cannot be used to infer haplotype the complex *CYP2A6* or *CYP2D6* loci due to concerns regarding sequence/alignment data quality, the presence of gene copy number variation and/or rearrangement with their respective pseudogenes.

Evidence level LIMITED

For haplotype/allele definitions lacking important information. This level is assigned to haplotypes defined by the P450 Nomenclature Committee before the transition to PharmVar that do not fulfil PharmVar criteria. For non-CYP genes, this evidence level may also be assigned if haplotypes were published using star allele designations before the gene was introduced to PharmVar and pertinent information supporting a haplotype is not available. Lastly, haplotypes with structural variations as described under 13b will also receive an evidence level of 'LIMITED' since their structures and/or breakpoints can only be approximated.

Haplotypes falling under this evidence level include the following examples:

- The gene has been screened across exons with a method other than sequencing, and regions suspected to contain 'mutations' were subsequently sequenced to determine the nature of the sequence variation. These alleles may carry sequence variation(s) in the regions that were not sequenced;
- Haplotypes for genes <10kb that were not entirely sequenced and lack information for upstream/downstream regions and/or introns;
- Haplotypes that have been computationally inferred based on weak evidence e.g. haplotype inferred in a small population (under 500 subjects); the haplotype was predicted to be present in a single subject only; the genotype or second allele of the subject with new haplotype was not specified.

Examples

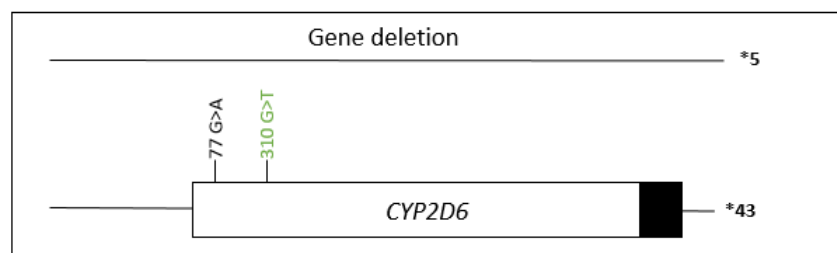
The following include three examples of submissions PharmVar has received to illustrate the criteria for allele/haplotype assignments, the value of 'encore' submissions and concept of evidence levels.

Submitters are encouraged to provide cartoons such as those shown below (in addition to the submission form) to facilitate the review of submissions, especially for complex haplotypes and those for which haplotypes are inferred through inheritance.

Example 1

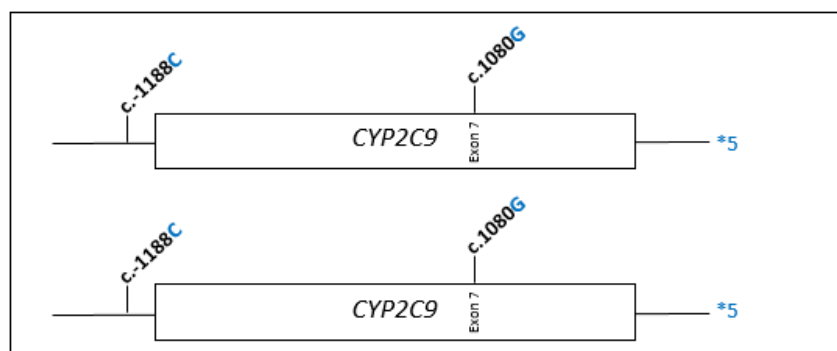
A subject was genotyped as *CYP2D6**5/*43 by sequencing a long-range (XL) PCR product encompassing the entire gene including about 1.7 kb of the upstream region. Since the subject carries a deletion of the entire gene (*5) on one allele, the XL PCR product was generated from one chromosome only and therefore both SNPs identified must be located on the same allele (as shown in the cartoon below). The genetic variation at position 77 signifies *43 while the one at position 310 (located within intron 1 of *CYP2D6*) has not been described for *43, although it has been found on numerous other alleles before. The 310G>T variant does not have any known functional consequence; the allele was therefore designated as a suballele of *43.

Since Sanger sequencing was performed and the haplotype was unequivocally determined, it received an evidence level of 'DEFINITIVE'.



Example 2

Whole Genome Sequencing (WGS) of Coriell sample NA23275 revealed genetic variations consistent with a *CYP2C9**5/*5 genotype. WGS fulfilled all quality requirements and the sample was homozygous for two variants, which unequivocally determines the haplotype. One variant (-1188T>C) was identified in the upstream region, which was not investigated by the authors first describing the *5 haplotype. This 'encore' submission allowed PharmVar to revise the *5 allele definition to include -1188T>C. This submission received an evidence level of 'DEFINITIVE' (subject was homozygous for all genetic variations detected). With this submission, the evidence level for *5 was raised from 'LIMITED' (due to missing information for the upstream region) to 'DEFINITIVE'.



Example 3

WGS was performed on a family trio. *CYP2C19* sequence information covered all exons, flanking introns and >3 kb of upstream region. *CYP2C19* genotypes were called as shown below in the pedigree.

This trio is informative for three haplotypes, *CYP2C19**1B, *17 and *35. Information was submitted to PharmVar for each of the three haplotypes. The graph below visualizes the two alleles of the mother.

Submission for *CYP2C1935:** Both mother and child carry variant -913G>A, a novel variant that is not part of any other haplotype definition. The graph on the left shows the variants on each of the mother's alleles as determined by inheritance. Since the father does not have -913G>A, it must be inherited via *35. The allele first defined as *35 (now referred to as *35.001) does not have -913G>A, hence the haplotype found in this trio was designated *35.002. The haplotype constitutes a suballele, because it is not known whether -913G>A alters function.

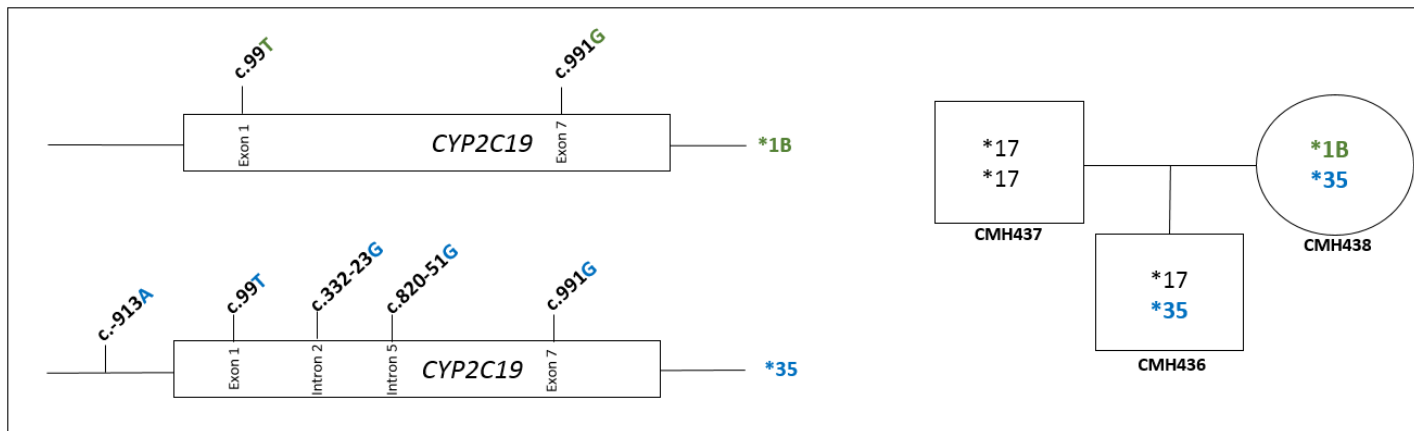
This submission for the novel *35 suballele received an evidence level of 'DEFINITIVE' (information for 1 pedigree with 2 subjects having the *35 haplotype).

Submission for *CYP2C1917:** The father is homozygous for all genetic variants shown. The *17 allele found in the father matched 100% with the original *17 definition.

This 'encore' submission received an evidence level of 'DEFINITIVE' (subject homozygous for all genetic variations detected; in addition, pedigree with 2 subjects having the *17 haplotype); this submission confirmed the *17 haplotype as originally defined.

Submission for *CYP2C191B:** The presence of *CYP2C19**1.002 (legacy label *CYP2C19**1B) was inferred through inheritance of the genetic variations detected. Two variants, c.99C>T and c.991A>G, were not passed on to the child via the *35 allele and therefore must be located on the mothers second allele, i.e. *1.002. This suballele was originally described in 1997; at this time sequencing did not cover the upstream region which is now known to harbor variants impacting function; it has therefore been listed with an evidence level of 'LIMITED'.

This 'encore' submission for *1.002 received an evidence level of 'DEFINITIVE' (information for 1 pedigree with 1 subject having the *1.002 haplotype) elevating the displayed evidence level for this allele from 'LIMITED' to DEFINITIVE'.



In this example, the ID numbers of the subjects in the Children's Mercy Genome Center database are provided.

For trios obtained from the Coriell Institute, respective sample ID numbers must be provided.