Important Information

DYPD is the rate limiting enzyme involved in the catabolism of fluoropyrimidines (5-fluorouracil and capecitabine) (1, 2). Clinical laboratories provide preemptive *DPYD* genotyping to avoid potentially life-threatening toxicity in patients carrying *DPYD* risk alleles. PharmVar is providing information describing allelic variation including intragenic deletions of one or more exons of this important pharmacogene. More information regarding *DPYD* copy number variation is provided in the Structural Variation document.

Prior to its curation in PharmVar, there was no centralized resource for *DPYD* nomenclature. Some of the allelic variants were assigned star allele numbers when first published (*DPYD*1-*13*) or were referred to by a trivial name (e.g., HapB3). Additionally, many important sequence variations were referred to by their rsID number or SNV position (e.g., c.2846A>T). Although initially proposed by McLeod and colleagues (3), the utilization of star nomenclature based on haplotypes built from the phasing of SNVs (in defined gene regions) was deemed impractical for *DPYD* due to the size of the gene (843,317 bp) and the presence of recombination between exons which makes haplotype phasing across all exonic regions extremely difficult. Furthermore, many of the consequential sequence variants are rare, and if detected, clinicians will likely act upon it regardless of the presence of other SNVs in the haplotype.

Therefore, to address these gene-specific challenges in nomenclature, PharmVar has developed a gene page format using rsIDs as PharmVar names to describe variants instead of star allele designations (rs.1 and rs.2 extensions are used to discriminate triallelic variants). Since intragenic deletions do not have rsIDs, PharmVar uses descriptors that include the affected exons. For example, “exon 4 del” denotes a “*DPYD* exon 4 deletion”.

Gene Region Mapped

PharmVar considers all nonsynonymous SNVs. **Intronic variants are only included if they are within splice recognition sequences, or there is convincing evidence that the variant impacts protein function.** Examples of listed intronic SNVs include c.1905+1C>T (rs3918290) and c.1129-5923C>G (rs75017182). See Allele Designation and Evidence Levels for additional details.

Coordinates

Variant coordinates in the PharmVar database are available for genomic and transcript reference sequences (RefSeqs) and the Human Genome Assemblies GRCh37 and GRCh38. **Table 1** summarizes important information regarding these sequences.

In this document coordinates are counted from the “A” in the ATG translation start codon as +1 for the NM_000110.4 as reference unless indicated otherwise.

Table 1

Sequence	notes
DPYD	Encoded on the negative strand.
NG_008807.2	Genomic reference sequence (RefSeq). NG_008807.2 is used as a reference to define allelic variation. PharmVar submissions must be annotated using this RefSeq.
LRG_722	NG_008807.2 matches the Locus Reference Genomic (LRG) sequence.
NM_000110.4	Transcript RefSeq. NM_000110.4 represents the MANE Select transcript that best represents gene function, is expressed, highly conserved and is well-supported by experimental evidence. MANE Select transcripts match GRCh38 and are identical to their genomic references for the 5'UTR, coding sequence, splicing, and 3'UTR,
GRCh37 and GRCh38	The *DPYD* LRG is identical to the locus sequence in both GRCh37 and GRCh38

PharmVar Feature: Get coordinates from the sequence start or the ATG start codon with one click! The default setting is NM_000110.4 counting the "A" in the ATG start codon as +1. The ⓘ buttons link out to NCBI to access respective sequences.

Sequence Variants

Sequence variants detailed in the PharmVar database comprise single nucleotide polymorphisms (SNPs) and insertions and deletions of a single or multiple nucleotides, collectively called single nucleotide variants (SNVs) or simply "variants". Structural variants include intragenic deletions which may affect one or more exons.

DPYD SNVs in the PharmVar database

All SNVs within the coding region have been published with *in vivo*, *in vitro* and/or *in silico* functional information. This set of SNVs is not intended to include all known SNVs.

There are several intronic SNVs which have been described in the literature to be in LD with variants that have been shown to associated with toxicity. Since these SNVs do not cause altered function themselves, or no information is available of whether they impact function, these are not listed by PharmVar.

The deep intronic risk variant c.1129-5923C>G (rs75017182) and the HapB3 haplotype

The haplotype known as HapB3 is defined by PharmVar as having two variants, c.1129-5923C>G (rs75017182), a deep intronic SNV that causes alternative splicing and results in decreased enzyme activity, and a benign synonymous SNV, c.1236G>A (rs56038477, p.Glu412=). These variants appeared to be in perfect linkage disequilibrium (LD) (see e.g., 1000 Genomes project data). Originally, the HapB3 haplotype was defined on the basis of a relatively small cancer patient cohort as the third most frequent haplotype of a haplotype block within *DPYD* and included three intronic variants, c.483+18C>T (rs56276561), c.680+139C>T (rs6668296) and c.959-51C>T (rs115632870) and c.1236G>A (4). The c.1129-5923C>G variant was added after it was discovered later to be also in LD with these variants (5, 6). Because c.483+18C>T is not in perfect LD with c.1129-5923C>G across populations, and c.680+139C>T is not exclusive to this haplotype, these two variants are not suitable proxies for c.1129-5923C>G (5). PharmVar does not display these intronic SNPs because they are not associated with altered function by themselves.

New data revealed that **c.1236G>A and c.1129-5923C>G are not in perfect LD as previously assumed**, i.e., c.1236G>A can, in rare cases, occur without c.1129-5923C>G (7). All identified cases were of European ancestry except for one which was “Other”. Consequently, testing strategies that only interrogate c.1236G>A may return a false-positive result predicting decreased function (intermediate metabolizer) in up to 0.2% of patients that may prompt inappropriate recommendations for dosing adjustments to prevent toxicity.

Laboratories may test the synonymous variant c.1236G>A, and not the intronic causal decreased function SNV c.1129-5923C>G to predict an individual’s risk of severe fluoropyrimidines-related toxicity. A test report may not specify that c.1236G>A is being used as a proxy variant to tag the functionally relevant c.1129-5923C>G and report c.1236G>A as the “risk” allele. The use of c.1236G>A as a proxy for the deep intronic c.1129-5923C>G is common practice when interpreting whole exome sequence data (WES) which does not cover the causal intronic variant.

For clarity, laboratories reporting risk based on c.1236 (p.Glu412=) genotype are advised to include appropriate language in their report stating that the causal variant c.1129-5923C>G may, in rare cases, not be present and thus prompt a false-positive result predicting decreased function (intermediate metabolizer).

***DPYD* testing strategies should be revised and/or updated to include the functionally relevant variant c.1129-5923C>G rather than rely on the c.1236G>A tag SNP to ensure DPD activity is accurately predicted while minimizing the risk of adverse events.**

PharmVar is now listing c.1129-5923C>G on its own and in combination with c.1236G>A (of which only the latter entry is cross-referenced with the “HapB3” as legacy name to acknowledge that c.1236G>A is being used as a tagSNP). Although c.1236G>A can occur without c.1129-5923C>G (5) it is not listed separately as such an entry would display “normal function”

and likely cause confusion with reporting when this SNV is used as tagSNP and interpreted as "decreased function" based on the assumption that c.1129-5923C>G is present.

Variant Window

Click on a variant to open this window. The example in **Figure 1** shows c.1129-5923C>A.

This view provides SNV positions across all sequences, links to external resources including dbSNP, PharmGKB and gnomAD, and frequency information.

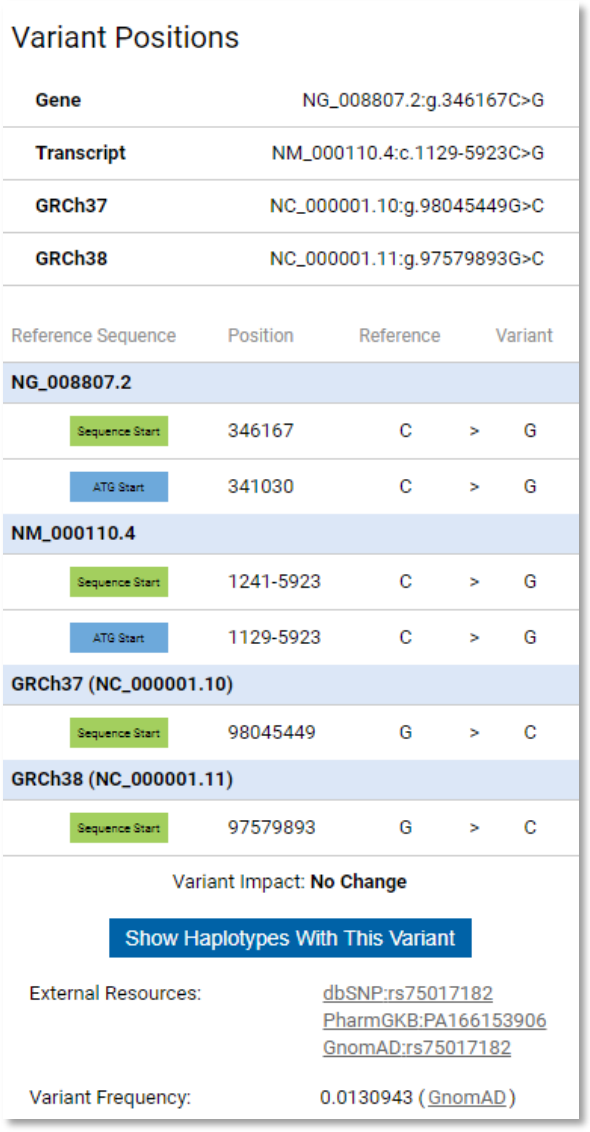
The **top portion** of the variation window **displays SNV coordinates according to Human Gene Variation Society (HGVS)** nomenclature at the gene, transcript and genome (GRCh37 and GRCh38) levels. Coordinates are displayed as obtained through the NCBI Variation Services and are only manually curated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

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

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 such variants which is most likely due to differences in sequence alignments. Also, PharmVar displays single nucleotide insertions as "ins" while HGVS displays them as duplications or "dup" if a base is duplicated. Additional details and examples are provided under Standards.

The **bottom portion** of the variant window provides easy access to dbSNP, PharmGKB and gnomAD allowing easy access to these external resources to retrieve additional frequency information for the variant besides its gnomAD global frequency which is displayed here.

Figure 1 Variant Window



Filter Options

A unique feature of the *DPYD* gene page is the column providing variant frequencies which can be sorted from highest to lowest or vice versa using the arrow in the column header.

Frequencies shown represent global frequencies sourced from the Genome Aggregation Database (gnomAD). See below for additional information on variant frequencies.

Frequencies of the two SNVs of the HapB3 haplotype are shown individually. Differences in their frequencies reflect that these are not in complete LD. Additional information regarding frequencies can also be found in Turner et al (5).

Additional filter options are available through the filter function [ ]. **SNVs can be filtered by function, frequency and position, or a combination thereof**.

Function

PharmVar displays "CPIC clinical allele function" as determined by the Clinical Pharmacogenetic Implementation Consortium (CPIC). The *DPYD* allele functionality table is available at PharmGKB https://www.pharmgkb.org/page/dpydRefMaterials. This table is created as part of the CPIC guideline process following the CPIC Pharmacogene Curation SOP, and is periodically updated. Additional information may have become available since the functionality table has been created.

Variant Frequencies

PharmVar displays the frequency of a sequence variation in the **Variant Window** using gnomAD as resource. The frequency shown represents the global frequency of the SNV which may considerably vary across populations.

Frequencies for the SNV can also be obtained by the external links to dbSNP, PharmGKB and gnomAD (see **Figure 1**).

Variant frequencies are also available through the *DPYD* allele frequency table at https://www.pharmgkb.org/page/dpydRefMaterials.

References

The references currently provided in the PharmVar database include publications primarily related to functional characterizations. These will be complemented by references including those first describing a SNV. For some alleles, additional reference(s) may be included in support of the original findings. The reference list is not intended to provide a complete bibliography for an allele. SNVs submitted to PharmVar but not published elsewhere are listed as "deposited by".

Changes and Corrections

Changes and edits are logged on the **CHANGE Log** document.

References cited in this Read Me document

LD between rs75017182 and rs56038477 calculated for all populations of the 1000 Genomes project using the LD Pair tool of the National Cancer Institute:
https://ldlink.nci.nih.gov/?var1=rs75017182&var2=rs56038477&pop=YRI%2BLWK%2BGWD%2BMSL%2BESN%2BASW%2BACB%2BMXL%2BPUR%2BCLM%2BPEL%2BCHB%2BJPT%2BCHS%2BCDX%2BKHV%2BCEU%2BTSI%2BFIN%2BGBR%2BIBS%2BGIH%2BPJL%2BBEB%2BSTU%2BITU&tab=ldpair

(1) Amstutz, U. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clinical pharmacology and therapeutics* **103**, 210-6 (2018) and updates at https://cpicpgx.org/guidelines/guideline-for-fluoropyrimidines-and-dpyd/

(2) Wei, X. *et al.* Characterization of the human dihydropyrimidine dehydrogenase gene. *Genomics* **51**, 391-400 (1998).

(3) McLeod, H.L. *et al.* Nomenclature for human DPYD alleles. *Pharmacogenetics* **8**, 455-9 (1998).

(4) Amstutz, U., Farese, S., Aebi, S. & Largiader, C.R. Dihydropyrimidine dehydrogenase gene variation and severe 5-fluorouracil toxicity: a haplotype assessment. *Pharmacogenomics* **10**, 931-44 (2009).

(5) Froehlich, T.K., Amstutz, U., Aebi, S., Joerger, M. & Largiader, C.R. Clinical importance of risk variants in the dihydropyrimidine dehydrogenase gene for the prediction of early-onset fluoropyrimidine toxicity. *Int J Cancer* **136**, 730-9 (2015).

(6) van Kuilenburg, A.B. *et al.* Intragenic deletions and a deep intronic mutation affecting pre-mRNA splicing in the dihydropyrimidine dehydrogenase gene as novel mechanisms causing 5-fluorouracil toxicity. *Hum Genet* **128**, 529-38 (2010).

(7) Turner, A.J. *et al.* Updated DPYD HapB3 haplotype structure and implications for pharmacogenomic testing. *Clin Transl Sci*, (2023).