Introduction

This document provides an overview of Structural Variations in the *CYP2D6* gene locus. Terminology and recommendations for structural variant (SV) annotations have been updated to be consistent with those described in the PharmVar tutorial on *CYP2D6* Structural Variation and Recommendations on Reporting (PMID 37669183) to standardize *CYP2D6* SV reporting.

Numerous structural variants (SVs) have been described, including gene deletions, duplications, and multiplications with two or more identical or non-identical gene copies, as well as structural rearrangements between *CYP2D6* and the *CYP2D7* pseudogene. SVs are often also referred to as copy number variants (CNVs) (see NCBI definitions for "Structural Variation"). While the term "SV" appears to be preferentially used in the literature when describing rearranged gene structures such as hybrid genes, the term "CNV" seems to be the preferred choice in the context of gene copy number testing. Since all known *CYP2D6* SVs are CNVs, they will be collectively referred to as SV/CNVs in this document. PharmVar-recommended terms and annotations are provided in **Appendix Tables A1 and A2**.

Table of Contents

CYP2D6 Reference Gene Locus

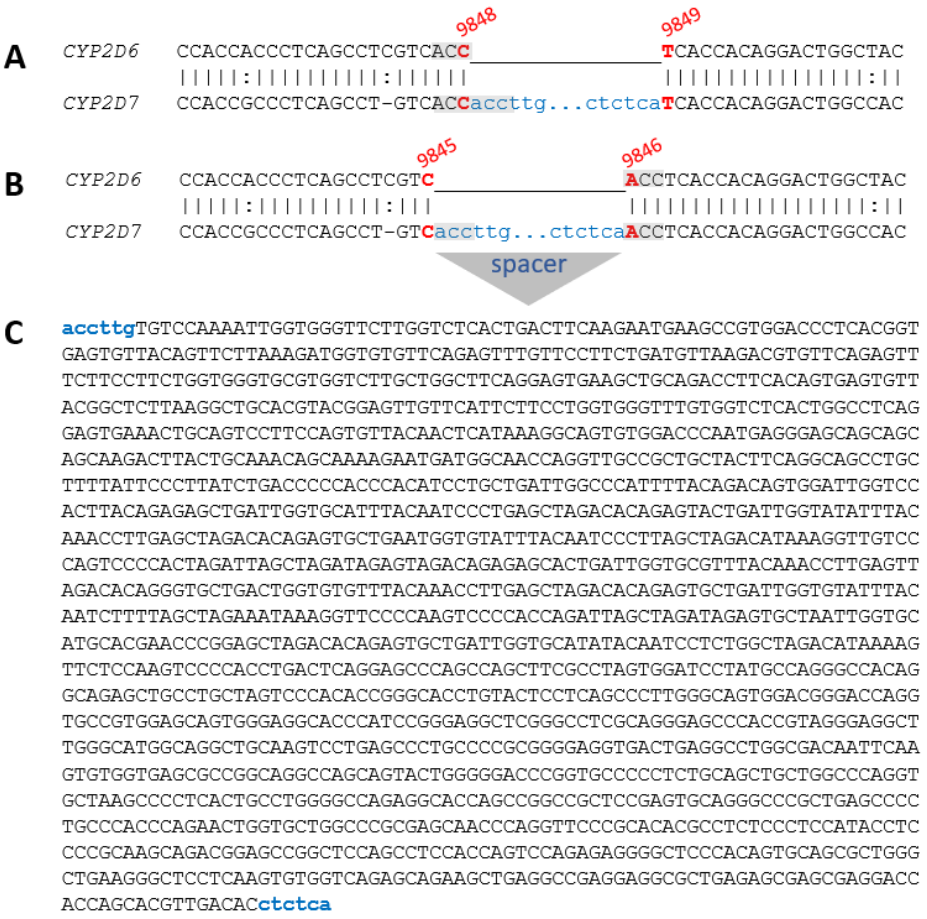
The gene locus contains three genes, *CYP2D6, CYP2D7,* and *CYP2D8* (**Figure 1**). All three genes are encoded by the negative strand (in reverse orientation) but are shown in forward orientation throughout this document. *CYP2D7* and *CYP2D8* are considered pseudogenes. All three genes encode nine exons and have high sequence similarity. *CYP2D6* and *CYP2D7* share a common 0.5 kb long downstream region (blue boxes) and have near-identical repetitive sequences referred to as REP6 and REP7, respectively.

A hallmark feature of the *CYP2D7* downstream region is a 1562 bp long "spacer" sequence that further distinguishes it from *CYP2D6*. We refer to *CYP2D6*-like and *CYP2D7*-like downstream regions, respectively, based on the absence or presence of the spacer region.

Figure 1 Reference *CYP2D6* Gene Locus

The spacer sequence can also be found downstream of some *CYP2D6* alleles. **Figure 2, panel A** depicts the insertion point within the *CYP2D6* downstream region on the genomic reference sequence NG_008376.4 (highlighted in red) using the 3' rule (the standard used by PharmVar for all alignments) which shows the insertion at the 3'-most position of the inserted sequence relative to the reference sequence. **Panel B** shows the insertion using the 5' rule showing the insertion at the 5'-most position. As can be seen, the alignments differ in positioning, with the insertion being after the "ACC" (3' rule) or before the "ACC" (5' rule). Notably, the first three bases of the spacer are also "ACC". The first and last six bases of the spacer sequence are shown in blue color and lowercase letters in the *CYP2D7* track of the sequence alignments in **panels A** and **B**, and the spacer sequence is shown in **panel C**.

Figure 2 The Spacer Sequence


A
CYP2D6 CCACCACCCTCAGCCTCGTCACC 9848 _____ 9849 TCACCACAGGACTGGCTAC
CYP2D7 CCACCGCCCTCAGCCT-GTCACCaccttg...ctctcaTCACCACAGGACTGGCCAC
B
CYP2D6 CCACCACCCTCAGCCTCGTC 9845 _____ 9846 ACCTCACCACAGGACTGGCTAC
CYP2D7 CCACCGCCCTCAGCCT-GTCaccttg...ctctcaACCTCACCACAGGACTGGCCAC
spacer
C
accttgTGTCCAAAATTGGTGGGTTCTTGGTCTCACTGACTTCAAGAATGAAGCCGTGGACCCTCACGGT
GAGTGTTACAGTTCTTAAAGATGGTGTGTTCAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGTT
TCTTCCTTCTGGTGGGTGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGTT
ACGGCTCTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTCACTGGCCTCAG
GAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAAGGCAGTGTGGACCCAATGAGGGAGCAGCAGC
AGCAAGACTTACTGCAAACAGCAAAAGAATGATGGCAACCAGGTTGCCGCTGCTACTTCAGGCAGCCTGC
TTTTATTCCCTTATCTGACCCCCACCCACATCCTGCTGATTGGCCCATTTTACAGACAGTGGATTGGTCC
ACTTACAGAGAGCTGATTGGTGCATTTACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTAC
AAACCTTGAGCTAGACACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCC
CAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTT
AGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTGGTGTATTTAC
AATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTAGATAGAGTGCTAATTGGTGC
ATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAG
TTCTCCAAGTCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAG
GCAGAGCTGCCTGCTAGTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGG
TGCCGTGGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGGGAGGCT
TGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGCGACAATTCAA
GTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCTCTGCAGCTGCTGGCCCAGGT
GCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCC
TGCCCACCCAGAACTGGTGCTGGCCCGCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTC
CCCGCAAGCAGACGGAGCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGG
CTGAAGGGCTCCTCAAGTGTGGTCAGAGCAGAAGCTGAGGCCGAGGAGGCGCTGAGAGCGAGCGAGGACC
ACCAGCACGTTGACACctctca

Sequences are aligned using the 3' rule:

NG_008376.4 position 9848 = CRCh38 position 42125960
NG_008376.4 position 9849 = CRCh38 position 42125959

Sequences are aligned using the 5' rule:

NG_008376.4 position 9845 = CRCh38 position 421259663
NG_008376.4 position 9846 = CRCh38 position 421259652

Variant *CYP2D6* Alleles

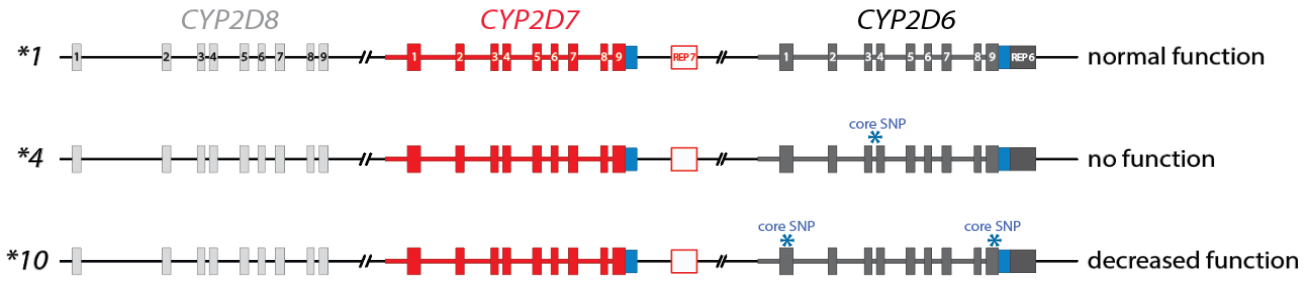
Alleles with SNVs and indels

Most *CYP2D6* allelic variants carry single nucleotide variants (SNVs), insertions, deletions, or combined deletion-insertion events (delins). Many alleles include one or more of these variants. In this document, for simplicity, these sequence variants are collectively referred to as single nucleotide variants (SNVs). As shown in **Figure 3**, alleles with SNV(s) encode proteins with normal, decreased, or no function (function remains unknown or uncertain for some). To date, the only alleles for which increased function has been established are those with gene

duplications. However, there is emerging evidence that *CYP2D6*53* may have higher activity than **1* (Muroi et al. 2014 [PMID: 24647041] and Glass et al. 2018 [PMID: 29784728]).

Not all known sequence variations within the *CYP2D6* gene locus are part of star alleles defined by PharmVar. There may be publications describing novel haplotypes that are not yet catalogued in PharmVar (i.e., no star allele designation). Investigators are encouraged to submit newly identified haplotypes to PharmVar for designation and dissemination to the pharmacogenetic community.

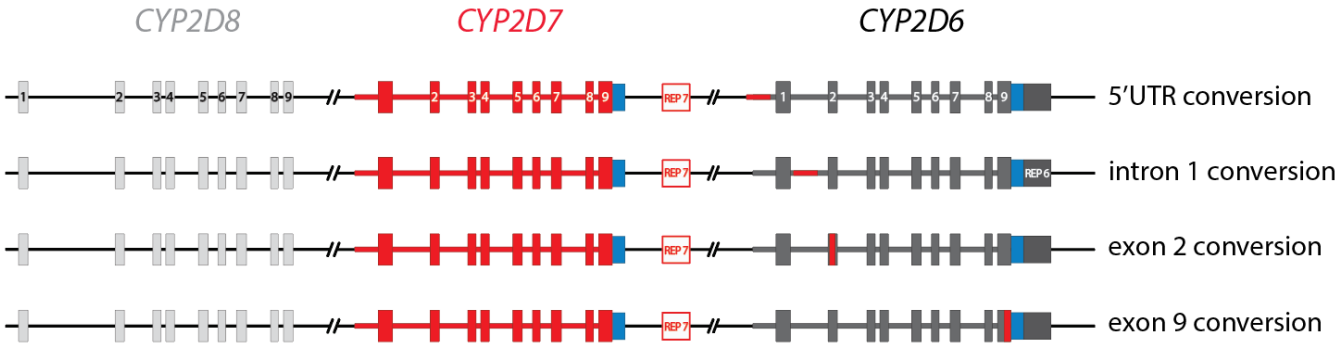
Figure 3 Examples of SNVs within the *CYP2D6* Gene
Blue asterisks denote core SNVs.



Conversions

Most *CYP2D6*2* suballeles have a *CYP2D7*-derived region within intron 1. This feature can also occur in other alleles and is often referred to as the "intron 1 conversion". Conversions have also been described for the 3'UTR, exon 2, and exon 9 (**Table 1**). These conversions are short *CYP2D7*-derived regions that are embedded within *CYP2D6* (**Figure 4**). The variants defining the 5'UTR, intron 1 and exon 2 conversions are listed along with other variants found in their respective haplotypes, but unlike the exon 9 conversion are not annotated as conversion. The

Figure 4 *CYP2D7*-derived Conversions within *CYP2D6*
The small conversion regions in *CYP2D6* are shown in red.



latter is annotated as a variation group (see the Read Me document for more detail). The intron 1 conversion of does not appear to affect function. The functional impact of the exon 2 conversion found in *CYP2D6*82* remains unknown. Amino acid changes associated with the exon 9 conversion likely contribute to a severe decrease in function if they do not render the allele completely nonfunctional. Note that alleles having the exon 9 conversion are category A hybrid genes (see page 9 below for more details regarding the classification of hybrids).

Table 1 Alleles and Suballeles Containing *CYP2D7* Conversions

Allele designation	*CYP2D7* conversions (positions per NG_008376.4, ATG start codon = +1)
****35.002***	**5'UTR conversion:** -431C>T, -354A>G, -334G>C, -331T>G, -328C>T, -327A>G, -321C>G, -320A>G, -276C>T, -275C>T, -272C>T, -268G>A, -267G>C, -232G>C, -225A>G
many alleles including **2*	**intron 1 conversion:** 214G>C, 221C>A, 223C>G, 227T>C, 232G>C, 233A>C, 245A>G for allele definitions based on exon sequence only, it remains unknown whether the intron 1 conversion is present. The number of alleles with the intron 1 conversion may therefore be underestimated.
****82***	**exon 2 conversion:** 973C>A (L91M), 983A>G (H94R), 996C>G, 1013T>C (V104A), 1021A>T+1022C>A (T107Y), 1027A>G (I109V), 1035T>C
****4.013, *4.031, *36, *83, *141***	**exon 9 conversion:** 4125G>C, 4129C>G (P469A), 4132A>G (T470A), 4134T>C, 4156C>T+4157A>C (H478S), 4159G>C (G479R), 4165T>G (F481V), 4167T>C, 4168G>A+4169C>G (A482S), 4170T>C, 4173C>T

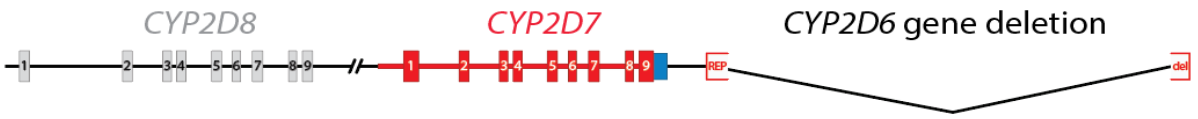
Copy Number Variation (CNV)

Gene Deletion

The allele defined as *CYP2D6*5* is characterized by a deletion of the entire *CYP2D6* gene with breakpoints in the *CYP2D7*-REP7 and the *CYP2D6*-REP6 regions, as shown in **Figure 5.** Thus, the downstream region of the *CYP2D6*5* allele is *CYP2D7*-like and contains the spacer and a REP region that has been described as REPdel. However, it is not known whether all deletion alleles identified as *CYP2D6*5* have identical breakpoints. ***CYP2D6*5* is described in the PharmVar database as "deletion of the entire gene".**

PharmVar recommends describing and reporting diplotypes with the *CYP2D6*5* gene deletion as **1/*5* or **4/*5* or as **5/*6* or **5/*29* (i.e., the star allele with the lower number is displayed first. For additional information on standardized reporting see **Appendix Table A2**.

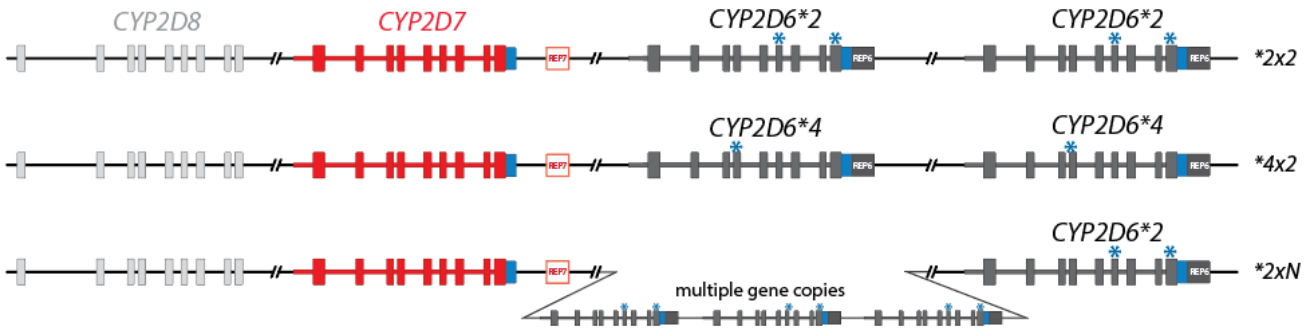
Figure 5 *CYP2D6* Gene Deletion



Identical Gene Duplications and Multiplications

Numerous *CYP2D6* alleles occur not only as singletons, but also in duplication or multiplication arrangements (**Figure 6**). Duplicated/multiplied gene copies have a *CYP2D6*-like downstream region without the spacer; their REP element is not identical to REP6 and is therefore referred to as REPdup. **Figure 6** provides selected examples of alleles carrying two identical gene copies (identical at the core allele level; these may differ at the suballele level). In very few instances were all gene copies sequenced. Gene copies are generally deemed identical, an assumption based on limited data (i.e., when both gene copies were genotyped and shown to carry the same set of core variants or, on rare occasions, sequenced). However, this may not always be the case. Also, duplicated alleles are more prevalent than allele multiplications.

Figure 6 *CYP2D6* Identical Gene Duplications



Identical duplications/multiplications have been described for many alleles, including those with normal function (e.g., *CYP2D6*1, CYP2D6*2*), decreased function (e.g., *CYP2D6*10, CYP2D6*41*), no function (e.g., *CYP2D6*4*), and uncertain function (e.g., *CYP2D6*43*). **Table 2** summarizes the gene duplications/multiplications that have been reported in the literature

and/or have been submitted to PharmVar. PharmVar does not list gene duplications as separate entities in the database.

If the gene copy number is known, duplications/multiplications are recommended to be annotated and reported as *x2, x3*, etc., and if the number is unknown as *xN*. If it is not known which of the two chromosomes carries the duplication and/or the number of gene copies is unknown, diplotypes should be described as detailed in **Appendix Table A2**.

To date, only *CYP2D6*1, *2, *4,* and **41* have been described to have 3 or more copies, whereas an increasing number of other star alleles have been described in the duplicated state (**Table 2**).

Table 2 *CYP2D6* Identical Gene Duplications/Multiplications

Gene copy number	CPIC clinical function	Activity Score value	References	PMID
1x2	increased function	2	Sachse et al. 1997 Gaedigk et al. 2007 Hosono et al. 2009 Gaedigk et al. 2012 Qiao et al. 2016 Del Tredici et al. 2018	9012401 17259947 19541866 22111604 26602992 29674966
1x ≥3	increased function	≥3	Gaedigk et al. 2012	22111604
2x2	increased function	2	Dahl et al. 1995 Aklillu et al. 1996 Gaedigk et al. 2007 Hosono et al. 2009 Gaedigk et al. 2012 Qiao et al. 2016 Del Tredici et al. 2018	7616439 8764380 17259947 19541866 22111604 26602992 29674966
2x ≥3	increased function	≥3	Johansson et al. 1993 Dahl et al. 1995 Aklillu et al. 1996 Gaedigk et al. 2012 Qiao et al. 2016	7903454 7616439 8764380 22111604 26602992
3x2	no function	0	Del Tredici et al.	29674966
4x2	no function	0	Løvlie et al. 1997 Sachse et al. 1998 Gaedigk et al. 2007 Gaedigk et al. 2012 Qiao et al. 2016 Del Tredici et al. 2018	9170153 10022755 17259947 22111604 26602992 29674966

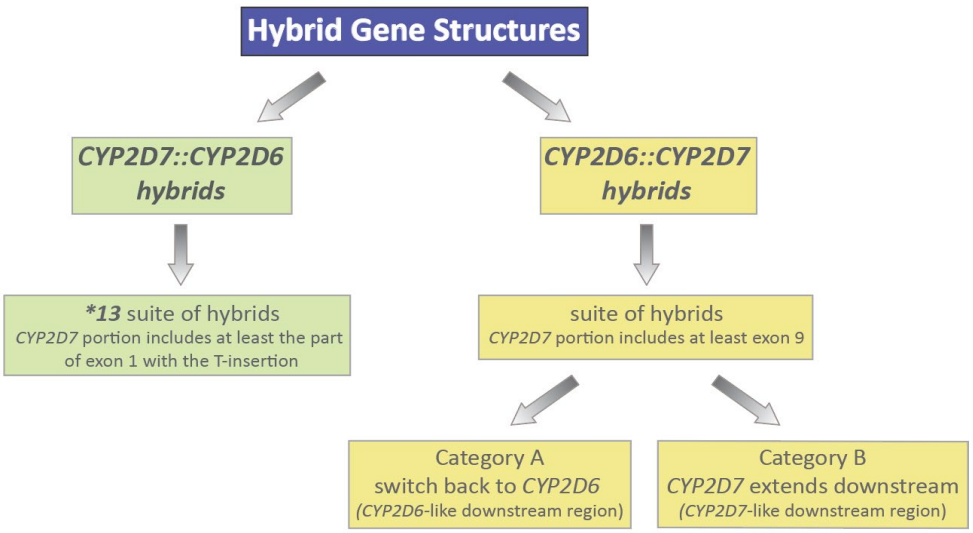
Gene copy number	CPIC clinical function	Activity Score value	References	PMID
4x≥3	no function	0	Qiao et al. 2016	26602992
4.013x2	no function	0	Gaedigk et al. 2012 Del Tredici et al. 2018	17259947 29674966
6x2	no function	0	Gaedigk et al. 2007 Del Tredici et al. 2018	17259947 29674966
9x2	decreased function	0.5	Gaedigk et al. 2011 Del Tredici et al. 2018	22044417 29674966
10x2	decreased function	0.5	Garcia-Barceló M, 2000 Ji et al. 2002 Mitsunaga et al. 2002 Ishiguro et al. 2004 Gaedigk et al. 2007 Hosono et al. 2009 Gaedigk et al. 2012 Del Tredici et al. 2018	10973875 12089164 12175908 15149890 17259947 19541866 22111604 29674966
17x2	normal function	1	Cai et al. 2006 Gaedigk et al. 2007 Gaedigk et al. 2012 Del Tredici et al. 2018	16550211 17259947 22111604 29674966
28x2 ***(*28.001+*28.003)***	uncertain function	n/a	Nofziger et al	PharmVar deposit
29x2	normal function	1	Gaedigk et al. 2007 Gaedigk et al. 2012 Del Tredici et al. 2018	17259947 22111604 29674966
35x2	increased function	2	Griese et al. 1998 Gaedigk et al. 2007 Gaedigk et al. 2012 Del Tredici et al. 2018	9511177 17259947 22111604 29674966
41x2	decreased function	0.5	Gaedigk et al. 2007 Gaedigk et al. 2012 Del Tredici et al. 2018 Gaedigk et al 201	17259947 22111604 29674966 31401124
41x3	decreased function	0.75	Gaedigk et al. 2007 Gaedigk et al. 2020	17259947 31401124
43x2	uncertain function	n/a	Gaedigk et al. 2007	17259947
45x2	increased function	2	Gaedigk et al. 2007	17259947
146x2	uncertain function	n/a	Gaedigk et al.	PharmVar deposit

Table 2 legend: Note: **10x2* has been revised to decreased function based on the CPIC recommendation for *CYP2D6* genotype to phenotype translation (Caudle et al. 2019, PMID 31647186). **9x2* and **41x3* have been revised to decreased function to reflect updated CPIC function assignments posted March 2023. Revisions are based on downgrading the value for activity score (AS) calculation for **9, *10* and **41* from 0.5 to 0.25. CPIC allele clinical function and AS values are according to the "allele functionality table" at https://www.pharmgkb.org/page/cyp2d6RefMaterials.

Hybrid Genes

Hybrid gene structures in *CYP2D6* include *CYP2D7::CYP2D6* hybrids, in which a 5' portion of *CYP2D6* including at least a portion of exon 1 has been replaced with *CYP2D7*, and *CYP2D6::CYP2D7* hybrids, in which the downstream region including at least exon 9 has been replaced with *CYP2D7* sequence. The symbols for hybrid genes are no longer separated by a hyphen or forward slash but by a double colon as recommended by HGVS. Adapting this recommendation for *CYP2D6* facilitates the immediate recognition of these special structures. **PharmVar classifies *CYP2D6::CYP2D7* hybrid genes as Category A or B based on the nature of their downstream region (Figure 7).**

Figure 7 Overview of Hybrid Gene Structure Classification



Category A alleles switch back to *CYP2D6* and have a *CYP2D6*-like downstream region (REP6) without the spacer. In contrast, Category B alleles have a *CYP2D7*-like downstream region with the 1.6 kb long spacer sequence (**Figure 7**). The latter have been described as REPdel because these structures are believed to be the result of unequal crossover events leading to deletion and duplication structures. **Figure 7** summarizes the different hybrid genes described to date. In rare cases, such as *CYP2D6*57*, the structure of the downstream region is unknown.

Some *CYP2D6::CYP2D7* hybrid genes may occur as singletons, in duplications, or both. The following sections describe the arrangements in which these hybrids have been found.

Table 3 *CYP2D7::CYP2D6* and *CYP2D6::2D7* Hybrid Genes

Allele	Hybrid structure	Switch region (legacy designations)	CPIC clinical function	References	PMID
***13**	***2D7::2D6***	intron 1 *(*13, *77)*, exon 2 *(*79)*, intron 2-exon 3 *(*80)*, intron 4 *(*78)*, exon 5 *(*67)*, exon 7 *(*66)*, intron 7-exon 8 *(*16)*[1], exon 9 *(*76)*,	no function	Panserat et al. 1995 Daly et al. 1996 Gaedigk & Coetsee, 2008 Gaedigk et al. 2010 Gaedigk et al. 2010 Black et al. 2012 Gaedigk & Turner[1] 2020	8554938 8873218 18202841 20017671 21833166 22004686 PharmVar deposit
***4.013**	***2D6::2D7***	within or upstream of exon 9; categories A and B	no function	Gaedigk et al. 2006	16415111
***36**	***2D6::2D7***	Within or upstream of exon 9; categories A and B	no function	Gaedigk et al. 2006 Hosono et al. 2009 Del Tredici et al. 2018	16415111 19541866 29674966
***61**	***2D6::2D7***	intron 7; category B (GenBank EU530607)	uncertain function	Kramer et al. 2009	19741566
***63**	***2D6::2D7***	exon 8; category B (GenBank EU530608)	uncertain function	Kramer et al. 2009	19741566
***68**	***2D6::2D7***	intron 1; category A	no function	Kramer et al. 2009 Gaedigk et al. 2012 Gaedigk 2020	19741566 22111604 PharmVar deposit
***83**	***2D6::2D7***	within or upstream of exon 9; category A	uncertain function	Gaedigk et al. 2012 Gaedigk et al. 2019 Nofziger 2020	22111604 31401124 PharmVar deposit

[1] switch in exon 7 between 3163 and 3255; maybe similar to that initially designated as **66* (also see **Figure 7**).

Legacy allele designations (those used prior to consolidation to **13* are shown in brackets).
CPIC allele clinical function and AS values are according to the "Allele functionality Table" available at https://www.pharmgkb.org/page/cyp2d6RefMaterials.

CYP2D7::CYP2D6 Hybrid Genes

CYP2D7::CYP2D6 hybrid genes have been grouped under a single star designation, *CYP2D6*13*. A hallmark feature of these hybrids is a *CYP2D7*-derived T-insertion in exon 1 causing a frameshift. **Since all *CYP2D6*13* hybrids have, by definition, a *CYP2D7*-derived exon 1 sequence containing the T-insertion, all of these hybrids are nonfunctional.**

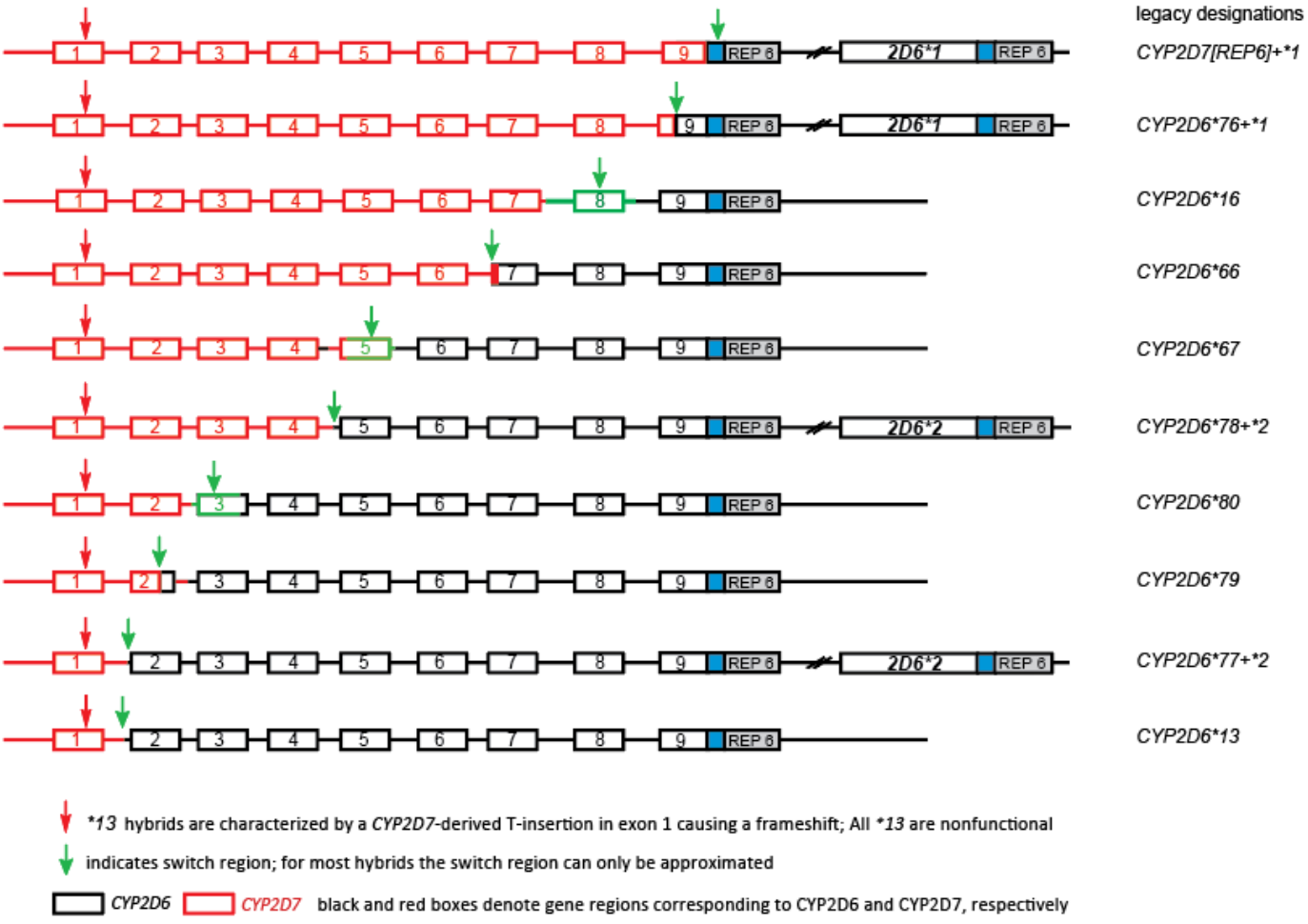
The 5' portion of these structural variants originates from *CYP2D7,* and the 3' portion is derived from *CYP2D6* (Figure 8). These hybrids are thought to be the products of large deletions between *CYP2D7* and *CYP2D6* and, to the best of current knowledge, thus lack an intact (complete) copy of *CYP2D7*. A number of hybrid genes have been reported that follow this structure, as summarized in **Figure 8** and **Table 3**. The hybrids have a *CYP2D6*-like downstream

region lacking the spacer. The allele shown in the top row of Figure 8 switches over past exon 9 and, thus, does not contain any *CYP2D6* coding regions. This structure is not a hybrid (see pages 11-13 for description of "*CYP2D6* genes with *CYP2D7*-like downstream regions").

Several *CYP2D7::CYP2D6* hybrid sequences have been deposited in GenBank: EU098008 and GQ162807 (switch region in intron 1), GQ162807 (switch region in exon 2), HM641840 (switch region in intron 2-exon 3), GQ162808 (switch region in intron 4), EU098009 (switch region in exon 5), JN618990 and HQ670229 (switch region in intron 7), EU093102 (switch region in exon 7-intron 8), and GQ162806 (switch region in exon 9).

***CYP2D6*13* hybrids are described in the PharmVar database as "*CYP2D7::CYP2D6* hybrid genes** (see Structural Variation Document for *CYP2D6*)".

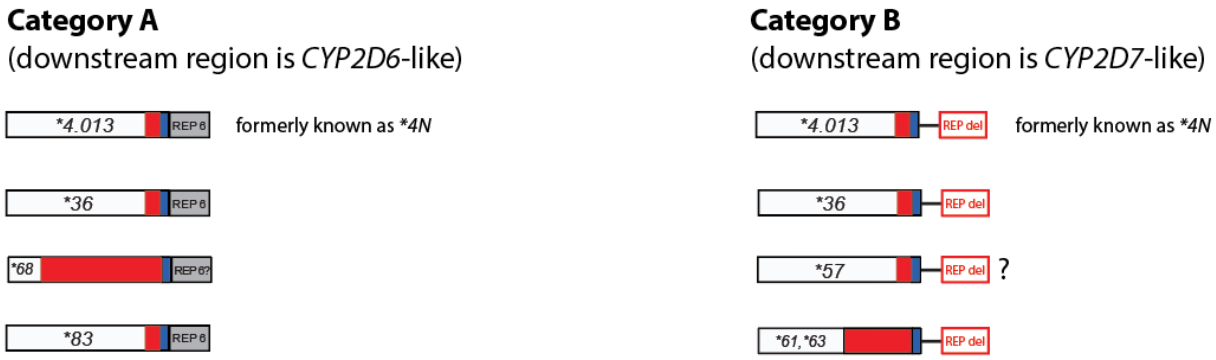
Figure 8 *CYP2D7::CYP2D6* (**13*) Hybrid Genes



CYP2D6::CYP2D7 Hybrid Genes

CYP2D6::CYP2D7 hybrid genes have not (yet) been consolidated under a single star designation because it remains unknown whether the exon 9 conversion, which is a common feature of these hybrids, drastically decreases or abolishes function. Categorization of these hybrids is also complicated because some switch back to *CYP2D6* after exon 9 (Category A), while the *CYP2D7* portion extends downstream in others (Category B) (**Figure 9**). Assays targeting this region may yield different results depending on which structure is present in the sample.

Figure 9 *CYP2D6::CYP2D7* Category A and B Hybrid Genes



CYP2D6::CYP2D7 Singletons

The following *CYP2D6::CYP2D7* hybrid genes were described as singletons: **36, *61, *63, *68,* and **83* (**Figure 10**). The most common singleton hybrid appears to be **36* of category A; **36.001* and **36.003* suballeles have been reported as singletons.

Of note, *CYP2D6::CYP2D7* singletons in category B (containing a REP7-like downstream region) may support amplification with certain XL-PCR-based *CYP2D6*5* assays, resulting in false-positive *CYP2D6*5* calls. More detailed information regarding XL-PCR amplification of SVs/CNVs can also be found in the PharmVar SV/CNV tutorial (PMID 37669183).

Additional *CYP2D6::CYP2D7* singletons are shown in **Figure 11** and **Table 4**.

Table 4 Singleton *CYP2D6* gene copies with a *CYP2D7* downstream region

allele designation	value for AS calculation	CPIC clinical function	references	PMID
***10.003**[REP7]*	0.25	decreased function	Ishiguro et al. 2004	15313161
***2.015**[REP7]*	1	normal function	Wang et al.	PharmVar deposit

CYP2D6::CYP2D7 in Duplication Arrangements

CYP2D6::CYP2D7 hybrid genes in duplication arrangements are category B hybrids (i.e., contain the "spacer" sequence followed by REP7) and are described in more detail below.

Figure 10 *CYP2D6::CYP2D7* Singleton Hybrid Genes

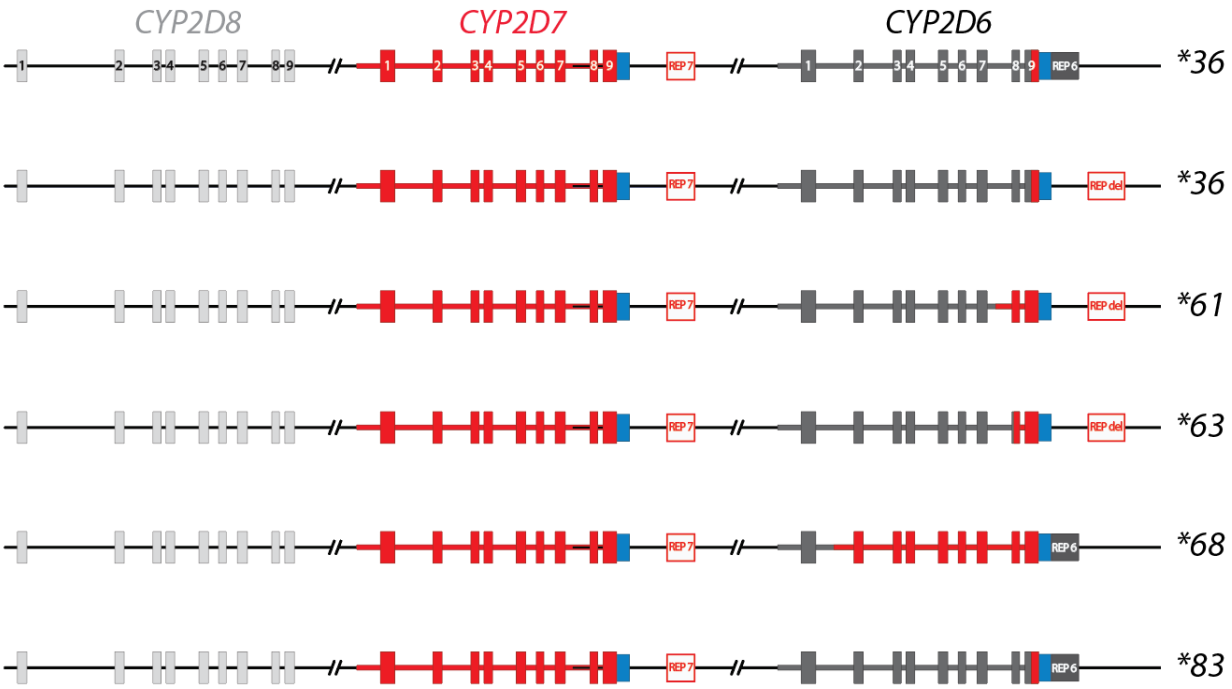
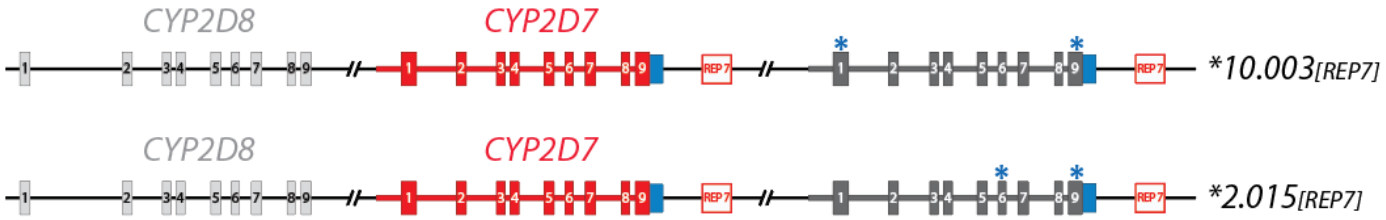


Figure 11 Singleton *CYP2D6* gene copies with a *CYP2D7* downstream region
Blue asterisks denote core variants.



Non-identical Gene Duplications

In the past, "tandem" was used to distinguish allelic variants with two or more gene units that are not identical from those with identical units that are duplicated or multiplied. PharmVar no longer recommends this term to avoid conflict with the NCBI's definition of "tandem" duplication. Such duplications are now called non-identical gene duplications (see **Appendix Table 2**).

As shown in **Figure 12**, non-identical duplications can harbor two or more gene copies. In most of these duplications, at least one gene copy is a *CYP2D7::CYP2D6* or *CYP2D6::CYP2D7* hybrid. The latter are category B hybrids containing a *CYP2D7*-derived downstream region, including the spacer.

PharmVar does not list non-identical gene duplications in the database. **Table 5** lists non-identical duplications that have been described in the literature and/or submitted to PharmVar.

One of the most frequently observed non-identical duplications is *CYP2D6*36+*10*, which is found primarily, but not exclusively, in individuals of East Asian ancestry, while the *CYP2D6*68+*4* non-identical duplication is frequently found in Europeans.

Interestingly**,** the most 3' gene copy in all non-identical gene duplications described to date has a *CYP2D6*-like downstream region (**Figure 12**).

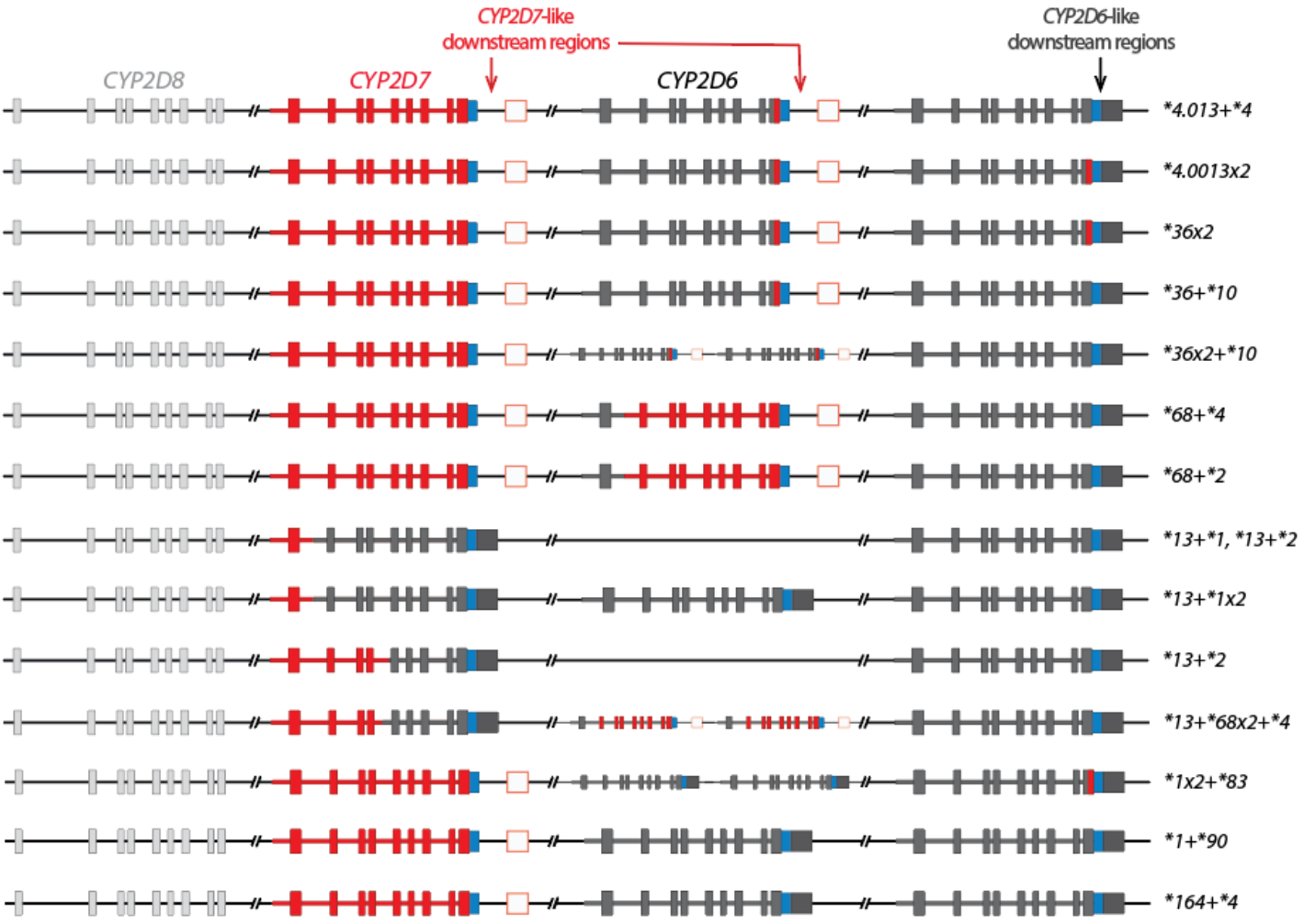
The 5' (upstream) gene copy of *CYP2D6*17x2* in most cases has a *CYP2D6*-like downstream region with a REPdup, as described above for other duplicated/multiplicated alleles (**Figure 6**). In rare cases, however, this gene copy has a *CYP2D7*-like downstream region with the spacer sequence and presumably a REP7 element, as shown in **Figure 12**.

The figure shows the most commmonly **13* hybrids found to date in duplication arrangements with a **1* or **2*; the switch from *CYP2D7* to *CYP2D6* may occur at regions different from those shown in **Figure 11**.

The *CYP2D6* portion of hybrid genes found in non-identical duplications often has the same sequence as the downstream *CYP2D6* gene. For example, the *CYP2D6* portion of the **36* in a **36+*10* duplication matches that of the **10* gene copy. However, this is not necessarily the case, as demonstrated by **36.004* and **36.*005 (the latter was originally designated as *CYP2D6*57*.

Lastly, the *CYP2D6*1x2+*83* and *CYP2D6*1+*90* duplications illustrate that the upstream (duplicated) gene copy (or copies) does not necessarily have a *CYP2D7*-like downstream region.

Figure 12 *CYP2D6* Non-identical Gene Duplications



*CYP2D6*68*: The *CYP2D6* portion of this hybrid encompasses exon 1 and most of intron 1. Several SNVs have been reported within the C*YP2D6* portion: -1426C>T, -1000G>A, 100C>T, 310G>T, 745C>G and 842T>G. The SNV at position 100C>T is the core SNV of *CYP2D6*10* (and other alleles) and is also found on most **4* alleles. While a heterozygous 100C/T call does not usually interfere with the interpretation of a genotyping result when a *CYP2D6*68+*4* is present, heterozygosity of 100C/T may be interpreted as 'inconsistent' (resulting in an indeterminate call or "no call") in the presence of a **68+*2* duplication and a non-**4* allele on the other chromosome.

Expert panel members involved in clinical testing have observed test results suggesting numerous additional SV/CNV structures with and without hybrid genes that have not been further characterized, published, or submitted to PharmVar.

Table 5 summarizes non-identical gene duplications that have been reported and/or submitted to PharmVar. Some of them represent complex structures with three or more gene copies. Their annotations reflect the order of the genes on the chromosome, i.e., *CYP2D6*36+*10* indicates that the **36* hybrid gene is upstream of the **10* gene copy, and *CYP2D6*68+*4* indicates that the **68* hybrid is upstream of the **4*. See **Appendix Table 2** for additional details and examples.

Table 5 Summary of Non-identical Duplications and Other Complex Structures

Allele designation	Activity Score value	CPIC clinical function	References	PMID
***1x2+*83**	2	increased function	Gaedigk et al. 2012	22111604
***4.013+*4**	0	no function	Del Tredici et al. 2018	29674966
***4.013xN+*4**	0	no function	Del Tredici et al. 2018	29674966
***4.013+*4xN**	0	no function	Del Tredici et al. 2018	29674966
***13+*1**	1	normal function	Gaedigk et al. 2010	20017671
***13+*1x2**	2	increased function	Black et al. 2012	22004686
***13+*2**	1	normal function	Gaedigk et al. 2010	20017671
***13+*68x2+*4**	0	no function	Black et al. 2012	22004686
***17**[REP7]**+*17** (suballele unspecified)	1	normal function	Gaedigk et al. 2012	22111604
***36x2**	0	no function	Gaedigk et al. 2006 Hosono et al. 2009 Del Tredici et al. 2018	16415111 19541866 29674966
***36x2+*10**	0.25	decreased function	Hosono et al. 2009 Gaedigk et al. 2012 Del Tredici et al. 2018	19541866 22111604 29674966

Allele designation	Activity Score value	CPIC clinical function	References	PMID
***36+*10**	0.25	decreased function	Johansson et al. 1994 Leathart et al. 1998 Gaedigk et al. 2006 Hosono et al. 2009 Gaedigk et al. 2012 Del Tredici et al. 2018	7935325 9918137 16415111 19541866 22111604 29674966
***36+*10.007**	0.25	decreased function	Wen et al. 2022	35387332
***36.004+*10.002**	0.25	decreased function	Wen et al. 2022	35387332
***36+*10x2**	0.5	decreased function	Hosono et al. 2009 Gaedigk et al. 2012 Del Tredici et al. 2018	19541866 22111604 29674966
***57+*10**	0.25	decreased function	Soyama et al. 2006	16858124
***68+*2**	1	normal function	Nofziger 2020	PharmVar deposit
***68+*4**	0	no function	Gaedigk et al. 2012	22111604
***1+*90**	(≥1)[1]	uncertain function	Gaedigk et al. 2019	31401124

[1] Function of **90* is uncertain and therefore, this allele has no value assigned to calculate Activity Score. However, since the **1* gene copy has normal function, this duplication is predicted to not have less than normal function.

See Table 2 footnote regarding the revised function of **10x2*.
CPIC allele clinical function and AS values are according to the "allele functionality table" at https://www.pharmgkb.org/page/cyp2d6RefMaterials.

References

The references provided in this Structural Variation document include the citation(s) in which an allele was first published. Additional reference(s) describe important updates and information regarding function. The reference list is not intended to provide a complete bibliography for an allele.

Users are encouraged to share their research with PharmVar and report important literature that might have been inadvertently overlooked.

Allele Frequencies

CYP2D6 allele frequency tables were developed for CPIC guidelines and are available through PharmGKB here. A list of frequencies, including population-specific information and references, can be found in the *CYP2D6* allele frequency table on the "references" tab. These tables are updated periodically.

Reference Materials

The Genetic Testing Reference Materials Coordination Program (GeT-RM) has developed materials for *CYP2D6* testing, including samples with structural variants, to facilitate copy number testing (PMID 31401124). A comprehensive list of samples can be found here.

Table 6 lists the samples reported by Get-RM that have a copy number variant. Samples can be obtained from the Coriell Institute for Medical Research (data accessed here April 25, 04-25-2023).

Table 6 Selection of Available Reference Materials for *CYP2D6* SV/CNV Testing

Structural variant	Coriell sample ID	*CYP2D6* diplotype
Gene deletion		
**5*	NA17235	**1/*5*
	NA18868	**2/*5*
	HG00276	**4/*5*
	NA18861	**5/*29*
	NA12336	**5/*41*
	HG00156	**5/*5*
	HG03225	**5/*56*
Identical gene duplications		
**1x2*	NA17454	**1x2/*2x2*
**2x2*	NA17454	**1x2/*2x2*
	NA19226	**1/*2x2*
	NA23296	**2x2/*4*
	NA19207	**2x2/*10*
	HG00337	**2x2/*22*
	NA19109	**2x2/*29*
**4x2*	NA19920	**1/*4x2*
	NA19819	**2/*4x2*
	NA15245	**4x2/*4*
	NA07439	**4x2/*41*

Structural variant	Coriell sample ID	*CYP2D6* diplotype
**4.013+*4*	NA10860	**1/*4.013+*4*
**10x2*	HG00421 HG00423 NA23246 NA23297	**2/*10x2* **10/*10x2* **10x2/*36+*10* **10x2/*17*
**17x2*	NA17113	**17x2/*45*
**36x2*	NA18565	**10/*36x2*
**41x3*	NA24217	**2/*41x3*
Hybrid genes (singletons)		
**83*	NA17287	**1/*83*
Non-identical gene duplications		
**1+*90*	NA18642	**36+*10/*1+*90*
**13+*2*	NA19785, NA19790	**1/*13+*2*
**36x2+*10*	NA18524, NA18526 NA18632	**1/*36x2+*10* **36x2+*10/*52*
**36+*10*	NA18563 NA18564, NA18959 HG00463, NA18617 NA23246 HG02373 NA18572	**1/*36+*10* **2/*36+*10* **36+*10/*36+*10* **10x2/*36+*10* **14/*36+*10* **36+*10/*41*
**36x2+*10x2*	NA18545	**5/*36x2+*10x2*
**68+*4*	HG01190 NA21781	*68+*4/*5 *2x2/*68+*4
No materials	**1x≥3, *2x≥3, *3x2, *4x≥3, *4.013x2,*6x2, *9x2, *29x2, *35x2, *41x2, *43x2, *45x2, *146x2, *13, *4.013, *36, *61, *63, *68, *1x2+*83, *4.013xN+*4, *4.013+*4xN, *13+*1, *13+*1x2, *13+*68x2+*4, *36+*10x2, *68+*2*	

Changes and Edits

Changes and edits are summarized in the **Change Log** document.

Appendix

Materials provided in this Appendix are sourced from the PharmVar tutorial on CYP2D6 Structural Variation and Recommendations on Reporting (CPT, accepted) and will be updated as necessary.

Table A1 Terms recommended by PharmVar to describe structural variation in pharmacogenes, including *CYP2D6*

Terms (abbreviation)	What the term means for *CYP2D6*
Star allele and haplotype	A star (*) allele describes a haplotype using a defined gene region (e.g., *CYP2D6*1, *2*). A star allele can have any number of variants and contain an SV/CNV. The region used to define *CYP2D6* star alleles encompasses 6066 bp (NG_008376.4 positions 3436 through 9501 counting from the sequence start or -1584 through 4482 counting from the translation start). The core allele for each star number represents all variants that i) lead to an amino acid change or ii) have been shown to alter function and are present in all suballeles. The term allele is often used to describe all variants present in a gene copy on the nucleotide level (e.g., a C to G change); *CYP2D6* star allele nomenclature should not be used to refer to a single variant unless that star allele only has a single variant (e.g., **18*). For example, 100C>T (rs1065852) itself should not be described as a *CYP2D6*10* allele or reported as *CYP2D6*10* allele frequency because the variant also occurs in many other star alleles or haplotypes, including the common *CYP2D6*4* allele.
Diplotype and genotype	A combination of two haplotypes (star alleles) constitutes the diplotype. For example, *CYP2D6*2/*10* indicates the presence of one **2* allele and one **10* allele. The term genotype is often used instead of diplotype, although the latter is more accurate.
Structural variation (SVs) and gene copy number variation (CNVs)	An SV is a rearrangement involving a region of DNA ≥1 kb. SVs include copy-neutral inversions and balanced translocations, as well as rearrangements resulting in genomic imbalances, such as insertions and deletions. These genomic imbalances are also commonly referred to as copy number variants (CNVs). All CNVs are SVs, but not all SVs are CNVs.

Copy number variation (CNV)	All CNVs are SVs. CNVs include gene deletions (copy number loss) and duplications or multiplications, with the latter representing copy number gains. A CNV may include one or more exons or whole genes and is larger than 1 kb. There are no known examples of intragenic duplications or deletions of only one or more exons of *CYP2D6*. The terms SV and CNV are often used interchangeably.
Multiallelic CNV (MCNV)	This term indicates that there are multiple forms of CNVs for a gene. *CYP2D6* can be described as a multiallelic CNV because there are alleles with deletions, duplications, or multiple gene copies. However, this term is rarely used to describe *CYP2D6* CNVs.
Insertion–deletion variants (indels and delins)	Insertion–deletion variants are DNA insertions, deletions, or combined insertion-deletion variants referred to as "delins". These are typically less than 1 kb. Although the term delins could also be used to describe small *CYP2D7*-derived regions within *CYP2D6*, these are typically referred to as "conversions" (see below).
Identical gene duplications	At least two identical or near-identical copies of the entire *CYP2D6* gene on the same allele, e.g., **1x2, *2x3, *4xN*. Over 90% of *CYP2D6* SVs involve identical copies of the gene. NCBI refers to a duplication of two identical, adjacent DNA regions as a tandem duplication. Gene duplications may be referred to as copy number gain in the literature or in reports.
Non-identical gene duplications	At least two non-identical gene copies on the same allele. Each gene copy has nine exons. One of the gene copies may be a hybrid with one or more exons corresponding to *CYP2D7*. Non-identical gene copies are defined as having different core star allele designations. Examples include **36+*10*, **68+*4*, **1+*90*, and **164+*4.* Non-identical gene copies should be listed in order of appearance, with the 5'-most gene copy listed first. Non-identical gene duplications have also been referred to as "tandem" duplications; however, to avoid conflict with NCBI's definition, PharmVar does not recommend this terminology.
Full/entire gene	Full or entire gene (or gene copy) indicates that all nine exons are from *CYP2D6* or *CYP2D7*.

Gene deletion	Deletion of the entire *CYP2D6* gene; this SV is designated as *CYP2D6*5.* Gene deletions may be referred to as copy number loss in the literature or in reports.
Hybrid gene	A gene copy consisting of nine exons containing portions of *CYP2D6* and *CYP2D7*; these are also referred to as "fusion" genes in the literature. Designations of *CYP2D7::CYP2D6* and *CYP2D6::CYP2D7* are used to distinguish which gene comprises the 5' portion of the hybrid. Examples include **13* (*CYP2D7::CYP2D6* hybrid) and **68* (*CYP2D6::CYP2D7* hybrid). PharmVar defines two categories of *CYP2D6::CYP2D7* hybrid genes. Category A hybrid genes have one or multiple *CYP2D7*-derived exon(s), including exon 9, but switch back to *CYP2D6* either late within exon 9 or the immediate 3'UTR; these hybrids have a *CYP2D6*-like downstream region and lack the *CYP2D7* spacer. In category B hybrid genes, the *CYP2D7* portion extends beyond exon 9; therefore, these entities have a *CYP2D7*-like downstream region. A **36* in a **36+*10* arrangement is typically a category B hybrid gene featuring a *CYP2D7*-like downstream region, while the singleton **36* in a **1/*36* represents a category A hybrid gene with a *CYP2D6*-like downstream region. Category A and B hybrid genes are not usually distinguished or specified in clinical reports.
***CYP2D7* conversions**	A small region of *CYP2D7*-derived sequence embedded within the *CYP2D6* gene. *CYP2D7* conversions in intron 1 and exon 9 are often referred to as "intron 1" and "exon 9" conversions. A singleton *CYP2D6*36* (category A hybrid) is an example of an allele with an embedded exon 9 conversion. Intron 1 conversion is found in several star alleles, including many **2* alleles. Conversions are typically not detailed in clinical test reports, where they may not be differentiated from hybrids.
Downstream gene copy	The 3'-most of two or more gene copies on the same chromosome. This gene copy may also be referred to as the non-duplicated gene copy. For example, *CYP2D6*10* is the downstream (3') gene copy of a **36+*10*. Of note, the non-duplicated copy may also refer to the allele on the opposite chromosome, e.g., the *CYP2D6*1* in a **1/*36+*10*.
Upstream gene copy or copies	The 5'-most gene copy (or copies) located between the *CYP2D7* gene and the downstream (3') gene copy. This gene copy may also be

	referred to as the duplicated gene copy or copies. For example, the *CYP2D6*36* and one of the **10* gene copies are the upstream (5') gene copies of a **36+*10x2*.
Copy Number (CN) assays	CN assays provide quantitative measurement of copy number at a given genomic location. For *CYP2D6*, the regions commonly used for CN assays include the 5'UTR, introns 2 and 6, and exon 9. CN assay calls at these loci are used to identify CNVs and infer the most likely gene structure.
Singleton	The term singleton is often used to indicate that a gene copy, specifically a hybrid gene such as *CYP2D6*36* or **68*, is not in a duplication arrangement but is the only gene copy present.

Table A2 PharmVar-recommended annotations and reporting of structural variants. The table provides a selection of SV/CNV alleles to exemplify the recommendations.

Recommendation	Examples	Interpretation
The allele with the lower star number is written first.	**1/*4*	A *CYP2D6*1* gene copy on one allele and a **4* gene copy on the other allele.
	**5/*10*	A full deletion on one allele and a *CYP2D6*10* gene copy on the other allele.
For identical gene copies on the same allele (in *cis*), the star allele is followed by an "x" and the number of gene copies. Gene copies may vary at the suballele level. **If the number of gene copies is unknown, write the star number followed by "xN".**	**1x2* **2x2* **4x2,* etc.	Two *CYP2D6* gene copies on the same allele.
	**2x3* **41x3*	Three *CYP2D6* gene copies on the same allele.
	**2xN* **36xN*	Two or more *CYP2D6* gene copies on the same allele.
When an SV/CNV is present, but the duplicated allele cannot be discriminated, write both possible diplotypes; i) if the gene copy number is unknown, write the star number followed by "xN"; ii) alternatively, the genotypes may also be reported in parentheses to denote ambiguity.	**2x2/*4* or **2/*4x2*	Two *CYP2D6*2* gene copies on one allele, and one **4* gene copy on the other allele, or one **2* gene copy on one allele, and two **4* gene copies on the other allele. This may introduce phenotypic ambiguity.
	**2xN/*4* or **2/*4xN*	The number of gene copies is unknown, and it was not determined which allele has the SV/CNV.
	*(*2/*4)x3* *(*2/*4)xN*	Provision of the genotype in brackets denotes that it was not determined which allele has the SV/CNV; x3 indicates a total number of three gene copies; xN denotes that the total number of gene copies is unknown.
For non-identical gene copies on the same allele (in *cis*), or multiple SVs/CNVs on the same allele, the	**1+*90*	One *CYP2D6*1* gene copy is upstream of one **90* gene copy on the same allele.

upstream gene copy is written first, followed by a "+" and the downstream gene copy. Although the order of the gene copies may not be experimentally determined in routine clinical testing, they should be displayed in their most likely order for consistency.	**68+*4*	One *CYP2D6*68* hybrid gene copy is upstream of one **4* gene copy on the same allele.
	**13+*2*	One *CYP2D6*13* hybrid gene copy is upstream of one **2* gene copy on the same allele.
	**36+*10*	One *CYP2D6*36* hybrid gene copy is upstream of one **10* gene copy on the same allele.
	**36x2+*10x2* **36+*10x2* **36x2+*10* **36xN+*10xN*	One or two *CYP2D6*36* hybrid gene copies are upstream of one or two **10* gene copies on the same allele; higher copy number states likely also exist.
	**13+*68x2+*4*	One *CYP2D6*13* hybrid gene copy is upstream of two *CYP2D6*68* hybrid gene copies upstream of one **4* gene copy, all on the same allele.
For diplotypes containing an SV/CNV, if both alleles have the same star number, write the allele without the SV first. For SV alleles with two non-identical gene copies, the star number of the downstream (3') gene copy determines whether the allele is displayed first or second in a diplotype.	**1/*1x2*	One *CYP2D6*1* gene copy on one allele and two **1* gene copies on the other allele.
	**2x2/*4*	Two *CYP2D6*2* gene copies on one allele and one **4* gene copy on the other allele.
	**68+*4/*10*	One *CYP2D6*68* hybrid gene copy upstream of one **4* gene copy on one allele, and one **10* gene copy on the other allele; the **68+*4* allele is written first due to **4* being the 3' gene with the lower star number.
	**2/*36+*10*	One *CYP2D6*2* gene copy on one allele and a **36* hybrid gene copy upstream of one **10* gene copy on the other allele; the **36+10* is written second due to **10* being the 3' gene with the higher star number.

	**1/*13+*2*	One *CYP2D6*1* gene copy on one allele, and a **13* hybrid gene copy upstream of one **2* gene copy on the other allele; the **13+*2* is written second due to **2* being the 3' gene with the higher star number.
For diplotypes that contain SVs/CNVs on both alleles, the downstream gene with the lowest star number is written first.	**2x2/*41x2*	Two **2* gene copies on one allele and two **41* gene copies on the other.
	**2x2/*68+*4*	Two *CYP2D6*2* gene copies on one allele, and one *CYP2D6*68* hybrid gene copy upstream of one **4* gene copy on the other allele.
	**13+*2/*36+*10*	One *CYP2D6*13* hybrid gene copy upstream of one **2* gene copy on one allele, and one **36* hybrid gene copy upstream of one **10* gene copy on the other allele.
	**68+*4/*1+*90*	One *CYP2D6*68* hybrid gene copy upstream of one **4* gene copy on one allele and one **1* gene copy upstream of one **90* gene copy on the other allele.
When an SV/CNV is present, but its structure has not been further validated, i) a default assignment may be used, or ii) all possible diplotypes can be written.	**13+*2/*4*	The *CYP2D6*13* gene copy is defaulted to being upstream of the **2* gene copy as **13+*4* has not been described. Other possible diplotypes include i) **2+*4/*13*, ii) **4+*2/*13*, or ii) **2*, **4*, and **13* on one allele and a **5* gene deletion on the other, although (i), (ii), and (iii) have not been described and are less likely.

For diplotypes with SV/CNV(s) and SNV(s) that cannot be reconciled with known star alleles and SV/CNV structures.	*n/a*	Samples that cannot be interpreted should be reported as 'indeterminate'. In such cases, it would be appropriate to inform the ordering provider that additional testing would be necessary to elucidate the SV/CNV structure(s) and determine the nature of the *CYP2D6* gene copies present.
Gene symbols for hybrids are separated by a double colon.	*CYP2D6::CYP2D7* or *CYP2D7::CYP2D6*	Gene symbols for hybrid alleles are no longer separated by a hyphen or forward slash but by a double semicolon per HUGO recommendation. Adapting this recommendation for *CYP2D6* facilitates instant recognition of such gene structures.