Legend to *CYP2D6* Allelic Variation Summary Table (R3)

A literature search was performed using the following search terms: CYP2D6, ethnic, ethnicity, populations, cytochrome P450 2D6, names of countries and/or populations such as 'Spain, Spanish, Brazil, Brazilian, etc or a combination thereof. In addition, reports were also identified from citations by others or review articles.

Manuscripts published before 1995 were excluded due to the sparse amount of genotype information provided in these reports. Also, only reports describing cohorts ≥50 subjects were included, with few selected exceptions (i.e. smaller cohorts were part of larger studies).

'-' indicates that this allele has not been tested. N/a indicates that no average, minimum or maximum allele frequencies are available. Healthy subjects and patient cohorts are listed separately or were combined in some instances if 'no difference' was reported. The number of total subjects was not determined for each major ethnic group, because some reports utilized the same cohort(s), or a substantial number of subjects, in more than one study, e.g. to determine additional sequence variations/allelic variants.

A few studies did not report allele frequencies, but genotype counts/frequencies. In these instances, allele frequencies were calculated from the data provided.

In some instances allele frequency calculation errors were detected and are shown corrected. In some instances authors were contacted for additional information and verification. Hence, allele frequencies in this table do not always correspond with those in their original publications.

*CYP2D6*1* was calculated as 100 - the sum of variants for a subset of reports. *CYP2D6*1* is indicated as '-', if a report did not provide data for **2* or only reported on a particular rare allelic variant or few variants which would dramatically over-estimate the frequency of *CYP2D6*1*.

For some alleles, including *CYP2D6*14* and **56*, manuscripts report an allele subtype. The reader is referred to the original article for further detail.

Note that frequencies tabulated for *CYP2D6*36* refer to the 'single' *CYP2D6*36* gene and not to the *CYP2D6*36+*10* tandem. This tandem is typically found in Asians and is often not discriminated from alleles carrying a 'single' *CYP2D6*10* gene. Particularly in Asians, reported *CYP2D6*10* allele frequencies comprise *CYP2D6*10* and the *CYP2D6*36+*10* tandem.

*CYP2D6*41* was tabulated as reported regardless of whether genotyping was performed using -1584C>G or 2988G>A. Note that the P450 AmpliChip Test determines *CYP2D6*41* via -1584C>G, which may particularly impact the accurate determination of *CYP2D6*2* and **41* in Africans and their descendants. *CYP2D6*41* allele frequencies may dramatically differ depending on whether -1584C>G or 2988G>A was used for its detection.

*CYP2D6*45* and **46* are shown separately and combined, because these alleles share a key SNP and are often not discriminated.

Note that some reports have used a re-sequencing approach. For some, only the alleles specifically reported on have been tabulated, while other alleles are shown as '-'. If absence of particular alleles could be inferred, their frequencies are shown as '0'. Please see original articles for specific information.

Many studies tested for the presence of 'gene duplications', but did not determine whether the duplication affected a *CYP2D6*1, *2, *4* or other gene. Often, duplications are defaulted to and

reported as *CYP2D6*2xN*. Unless duplications were clearly discriminated, they were tabulated in this summary table as 'undefined/other duplications'. Sometimes it may be possible to infer the nature of a duplication allele from a genotype (*e.g. CYP2D6*2/*2x2*); this however, could not systematically be captured in this tabulation. Also, some manuscripts reported on the number of gene copies, e.g. *x2, x3*, etc. Since there are few reports, these were combined with their respective *x2* group.

Some alleles, including *CYP2D6*1* and **2*, have a wide range of allele frequencies. These alleles are 'default' alleles, which means that if no sequence variations are detected or only a small number of variants are tested, an allele is assigned as (or defaulted to) *CYP2D6*1* or **2*. In such instances, *CYP2D6*1* and **2* frequencies are likely over-estimated.

The nomenclature for *CYP2D7/2D6* hybrid alleles has recently been updated. All alleles carrying a gene composed of *CYP2D7* and *2D6* and have a T-insertion in exon 1, have been consolidated under the *CYP2D6*13* designation. This change affects *CYP2D6*13, *16, *66, *67, *79* and **80*; each is shown in its own respective column, but are also shown as *'*13* revised'. Note that *CYP2D6*76, *77* and **78* have only been reported in tandem arrangements to date. These are listed separately along their revised genotypes.

This *CYP2D6* allele frequency summary may not be complete. If you wish to have your paper and/or data included into this resource, or identify any errors, please contact Andrea Gaedigk, PhD at agaedigk@cmh. This table will be periodically updated to include additional reports and *CYP2D6* allelic variants.

To access the *CYP2D6* allelic variation summary table in excel format please see http://www.pharmgkb.org/download.action?filename=CYP2D6_allele_frequency_table_R3_posted.xlsx

The excel table was last updated December 2014.