Frequency of undetected *CYP2D6* hybrid genes in clinical samples: Impact on phenotype prediction.

John Logan Black III, Denise L. Walker, Dennis J. O’Kane, Maria Harmandayan

Nucleotide Polymorphism Laboratory (JLB, DJO), Department of Laboratory Medicine and Pathology, and Psychogenomics Laboratory (JLB, DLW, MH), Department of Psychiatry and Psychology, Mayo Clinic and Mayo Medical School, Rochester, MN, USA 55902
Running Title: Frequency of undetected CYP2D6 hybrid genes in clinical samples

Address correspondence to: John Logan Black III MD, Department of Laboratory Medicine and Pathology, Mayo Clinic 200 1st Street SW, Rochester, Minnesota, USA 55902, phone: 507-284-3374, Fax: 507-266-4176, E-mail: black.john@mayo.edu.

Text pages: 39

Tables: 4

Figures: 3

References: 26

Abstract: 236 words

Introduction 746 words

Discussion 898 words

List of nonstandard abbreviations: None
Abstract

Cytochrome P450 2D6, (CYP2D6) is highly polymorphic. CYP2D6-2D7 hybrid genes can be present in samples containing CYP2D6*4 and CYP2D6*10 alleles. CYP2D7-2D6 hybrid genes can be present in samples with duplication signals and in samples with homozygous genotyping results. The frequency of hybrid genes in clinical samples is unknown. 1390 samples were evaluated for undetected hybrid genes by PCR amplification, PCR fragment analysis, Taqman® copy number assays, DNA sequencing and allele specific primer extension assay (ASPE). Of 508 CYP2D6*4-containing samples, 109 (21.5%) harbored CYP2D6*68 + *4-like, while 9 (1.8%) harbored CYP2D6*4N + *4-like. Of 209 CYP2D6*10-containing samples, 44 (21.1%) were found to have CYP2D6*36 + *10. Of 332 homozygous samples, 4 (1.2%) harbored a single CYP2D7-2D6 hybrid and of 341 samples with duplication signals, 25 (7.3%) harbored an undetected CYP2D7-2D6 hybrid. Phenotype before and after accurate genotyping was predicted using a method in clinical use. The presence of hybrid genes had no effect on the phenotype prediction of CYP2D6*4-and CYP2D6*10-containing samples. Four of 4 (100%) homozygous samples containing a CYP2D7-2D6 gene had a change in predicted phenotype and 23 of 25 (92%) samples with a duplication signal and a CYP2D7-2D6 gene had a change in predicted phenotype. Four novel genes were identified (CYP2D6*13A1 variants 1 and 2, CYP2D6*13G1 and a CYP2D6*13G2) and two novel hybrid tandem structures consisting of CYP2D6*13B + *68 X 2 + *4-like and CYP2D6*13A1 var 2 + *1 X N were observed.
Introduction

Cytochrome P450 2D6 (CYP2D6) encodes an enzyme clinically relevant to the use of tamoxifen (Goetz et al., 2005; Goetz et al., 2007) and the selection of psychiatric drugs (Black 3rd et al., 2007). CYP2D6 testing is sometimes done before tamoxifen treatment of breast cancer because patients who fail to metabolize tamoxifen to endoxifen via the CYP2D6 pathway may be at increased risk of breast cancer relapse (Goetz et al., 2007). Genotyping is also used to select psychotropic medications (Topic et al., 2000; Murphy et al., 2003; Kirchheiner et al., 2004; Malhotra et al., 2004).

The CYP2D locus contains the CYP2D8 and CYP2D7 pseudogenes 5’ to the CYP2D6 gene (Figure 1A). CYP2D6 genetic complexity is due to single nucleotide polymorphisms (SNPs), duplications and multiplications (Figure 1B), deletions (Figure 1C), and recombination events with the CYP2D7 pseudogene (Figure 2A-F). The allelic nomenclature of CYP2D6 is based upon the “star” (e.g. *) nomenclature method whereby alleles are named sequentially by the Allele Nomenclature for Cytochrome P450 Enzymes Committee (http://www.cypalleles.ki.se/index.htm). Subfamilies of alleles are given the same number but members of the family are given different suffixes. For example, CYP2D6*4 is a subfamily with A-N members and each has different SNPs associated with the defining SNP for CYP2D6*4; 1846G>A. CYP2D6*4N has a gene conversion to CYP2D7 in exon 9 while CYP2D6*4-like, which is mentioned in this article, is different from CYP2D6*4N by the fact that it lacks a conversion to CYP2D7 sequence in exon 9 (Kramer et al., 2009). Genotyping platforms are not designed to detect the exact allele present in the CYP2D6*4 subfamily so the suffix of A-N is often omitted in clinical genotyping results.
The recombinant events that occur with the CYP2D7 pseudogene may give rise to hybrid genes which occur singly or in hybrid tandem arrangements. Single hybrids can be of the CYP2D6-2D7 variety such as CYP2D6*61 which switches to CYP2D7 sequence from intron 7 onwards. Alternatively, single hybrids can be of the CYP2D7-2D6 variety. The Home Page of the Human Cytochrome P450 (CYP) Allele Nomenclature Committee has reclassified CYP2D7-2D6 hybrid genes that contain a CYP2D7-derived exon 1 (see http://www.cypalleles.ki.se/cyp2d6.htm for details) and this manuscript uses this new classification for the first time (personal communication – Sarah C. Sim PhD, Karolinska Institute, Stockholm, Sweden). A key to conversion to the new nomenclature is described in the Methods Section. An example of a CYP2D7-2D6 hybrid is CYP2D6*13A1 in which the switch from CYP2D7 to CYP2D6 sequence occurs in intron 1 (Figure 2A).

Hybrid genes can also occur in tandem arrangements. We recently described the CYP2D6*68 + *4-like and CYP2D6*4N + *4-like hybrid tandem arrangements (Kramer et al., 2009) while others described the CYP2D6*36 + *10 (Chida et al., 2002; Gaedigk et al., 2006). These three structures are of the CYP2D6-2D7 + CYP2D6 hybrid tandem variety (Figure 2F). Another hybrid tandem arrangement is the CYP2D7-2D6 + CYP2D6 variety (Figure 2B), an example of which is CYP2D6*13A2 [EU530609] + *2A (Kramer et al., 2009; Gaedigk et al., 2010a).

The aim of this study was to determine the frequency of hybrid genes present in a large number of samples originally genotyped using the Luminex Tag-It Mutation Detection Kit for Cytochrome P450-2D6 (CYP2D6 ASPE kit) and to identify the impact of these variants on phenotype prediction. The CYP2D6 ASPE kit does not detect hybrid genes because the kit generates A and B amplicons for the 5’ and 3’ portions of the CYP2D6 gene, respectively. Amplicon A mix also contains additional primers that allow for the amplification of rep dup
(Figure 1B) and its subsequent detection by an ASPE primer. In the case of CYP2D7-2D6 + CYP2D6, a duplication signal is detected even though a hybrid gene is present in the 5’ position (Figure 2B and 2C). CYP2D6-2D7 and CYP2D7-2D6 genes will cause failure of A and/or B amplicon generation during the PCR step depending upon the location of the switch from CYP2D6 to CYP2D7 sequence. Because there is a tendency for the CYP2D6 portions in the hybrid tandem gene to be identical to the CYP2D6 portion in the normal gene (Kramer et al., 2009), the kit will not detect partial failures in the ASPE step. In the case of a single CYP2D7-2D6 gene, total failure of the PCR of the hybrid gene can occur which can lead to a homozygous call based upon amplification of the non-hybrid allele. In this research the samples studied included: 1) CYP2D6*4-containing samples, 2) CYP2D6*10-containing samples, 3) samples with homozygous genotypes and 4) samples with duplication signal.

Materials & Methods

Samples and initial clinical genotyping

1390 de-identified samples submitted to the Nucleotide Polymorphism Laboratory of the Department of Laboratory Medicine and Pathology at the Mayo Clinic for clinical CYP2D6 genotyping using the Luminex Tag-It Mutation Detection Kit for Cytochrome P450-2D6 version 1 or 2 (hereafter named CYP2D6 ASPE kit v1 or v2, respectively) (Austin, Texas, USA) were studied with Mayo Clinic Institutional Review Board approval. The samples were submitted from practices both internal and external to the Mayo Clinic system but all were submitted either to predict tolerability and response to tamoxifen or psychotropic drugs. Samples are submitted from all over the world and because they were de-identified, the ethnic and racial mixture of these samples could not be determined. To meet criteria for study, a sample needed to have 1) at
least one CYP2D6*4 allele (N=508), 2) at least one CYP2D6*10 allele (N=209), 3) a homozygous genotype (N=332), or 4) a duplication signal (N=341).

The CYP2D6 ASPE kit v2 detects the following CYP2D6 recombinants and alleles: gene duplication, gene deletion (*5), – 1584 C>G (*2A), 100 C>T (*4, *10), 124G>A (*12), 138insT (*15), 883G>C (*11), 1023C>T(*17), 1661G>C (*2, *4, *17, and others), 1707T>del (*6), 1758G>T/A (*8, *14), 1846G>A(*4), 2549A>del (*3), 2613-2615 del AGA (*9), 2850C>T (*2, *17, *41 and others), 2935A>C (*7), 2988G>A (*41), 4180G>C (*2, *4, *17, others). The CYP2D6 ASPE kit v1 lacked the 138insT, 1661G>C, 1758G>A, 2988G>A, and 4180G>C polymorphisms. All samples originally genotyped with the v1 kits which contained a CYP2D6*2 allele with duplication were retested using the v2 kit in order to identify CYP2D6*41 alleles. The 138insT and 1758G>A have a very low frequency, as we have found 4 and 2 alleles, respectively in our experience of over 11,000 samples so we did not retest our samples for these. The 1661G>C and 4180G>C polymorphisms are part of several alleles and were added to the CYP2D6 ASPE kit v2 to help confirm allelic calls so retesting for these was not done. In 240 of 285 heterozygous duplication samples, the duplicated allele was predicted on the basis of the mutant allelic ratios as established for the clinical assay.

DNA extraction

DNA was extracted from blood using the DNeasy Blood and Tissue kit (Qiagen, Germantown, MD, USA) using the method described in the product literature.

PCR and amplicon verification

CYP2D6-2D7 and CYP2D7-2D6 genes were detected by PCR using primers specifically designed to allow them to be amplified as described in (Kramer et al., 2009) (also see Table 1 and Figure 1 and 2). The PCR primers used in this study were selected using Oligo Primer
Analysis Software version 6.71 (Molecular Biology Insights Inc., Cascade, Colorado, USA).

PCR amplicon locations are shown in Figure 1 and 2 as black bars under each structure. Specific primers were designed to prevent unintentional nonspecific annealing between the *CYP2D6* gene, *CYP2D7* pseudogene and *CYP2D8* pseudogene. All samples were screened for *CYP2D6-2D7* genes using amplicon E and *CYP2D7-2D6* genes using amplicon G regardless of the original genotype.

The PCR master mix was composed of 12 µL reactions containing 0.12 µL TaKaRa LA Taq HS (5 U/µL, TaKaRa Bio. Inc., Otsu, Shiga, Japan), 1.2 µL 10X LA PCR Buffer II (25 mmol/L Mg^{2+}), 2 µL dNTP mixture (2.5 mmol/L each), 6.48 µL water and 1 µL betaine monohydrate (16.25 mol/L, Fluka Biochemical, Steinheim, Germany), 0.5 µL of each primer and 0.2 µL genomic DNA (250 ng/µL).

Amplicons A-E, G, and J (Table 1) were generated as described by Kramer et al. (Kramer et al., 2009). Thermocycler parameters for amplicon K were as follows: 94°C for 1 minute followed by 96°C for 10 seconds, 64°C for 30 seconds, 68°C for 11 minutes for 30 cycles and a final extension at 72°C for 10 min and a 4°C hold. Thermocycler parameters for fragment L were: 94°C for 1 minute followed by 35 cycles of 96°C for 10 seconds and 68°C for 4 minutes, a final extension at 72°C for 10 min and a 4°C hold. Thermocycler parameters for fragment M and N were: 94°C for 1 minute followed by 30 cycles of 96°C for 10 seconds and 68°C for 9 minutes, a final extension at 72°C for 10 min and a 4°C hold.

PCR fragment sizes were analyzed using the Agilent Technologies DNA 12000 Kit (Agilent Technologies, Waldbronn, Germany) as per manufacturer’s protocol. Expected fragment sizes are shown in Table 1.
Genotyping by allele-specific primer extension assay

Genotyping of amplicons A-E, G, and J-M (Table 1) was performed using the Luminex Tag-It Mutation Detection Kit for Cytochrome P450-2D6 version 2 (CYP2D6 ASPE kit) (Austin, Texas, USA) as described in the product literature but without the original amplification step using the A and B primer sets. All samples were analyzed on the Luminex 100 IS device using IS 2.3 software (Austin, Texas, USA).

DNA sequencing

DNA sequencing was performed using an Applied Biosystems 3730xl DNA Analyzer and the ABI PRISM dRhoamine Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA Polymerase, FS. All CYP2D7-2D6 genes (amplicon G) and their associated rep dup regions (amplicon L) were sequenced bidirectionally using multiple overlapping amplicons. In addition, a unique structure containing a CYP2D6*13B + *68 X 2 + *4-like was discovered (described below) and was sequenced from exon 9 of CYP2D6*13B through exon 9 of CYP2D6*68 using amplicons E, G and K. Source of DNA for all sequencing was genomic DNA isolated from blood and amplicons were generated per PCR section noted previously.

Real Time PCR CYP2D6 copy number assays

The CYP2D6 copy number in samples containing both a CYP2D7-2D6 gene and a duplication signal was determined using two real time PCR (rtPCR) assays which targeted different regions of the gene (Figure 1 and 2). This assay was done only on the duplication samples because accurate copy number determination was needed to predict phenotype. Phenotype prediction for CYP2D6*4 and CYP2D6*10 containing samples is not impacted by copy number as will be discussed below so copy number was not determined for these samples. One assay,
Hs04502391_cn (intron 6) was commercially available (Applied Biosystems, Inc. (ABI), Foster City, CA). The other assay targeted the 5’-flanking region as described by Hosono et al (Hosono et al., 2009) was synthesized and purchased from ABI. Both assays were performed in triplicate with an internal control-RNaseP TaqMan® copy number reference and TaqMan® Genotyping PCR Master Mix on a StepOne Plus rtPCR instrument as directed by the manufacturer (ABI). Relative quantification was performed using CopyCaller® Software (ABI) following the comparative ΔΔC_T method. The confidence estimate for a given sample was automatically generated. It is the probability that the calculated copy number is the correct assignment compared to other copy numbers that have nonzero probability of occurring. The CopyCaller™ Software uses the copy number and observed ΔC_T values of a sample to calculate the confidence estimate.

**Bioinformatics tools**

DNA sequence data was compiled using Sequencher software, version 4.9 (Gene Codes Corp, Ann Arbor, Michigan, USA) and Mutation Surveyor (v3.13) (SoftGenetics LLC, State College, Pennsylvania). GenBank entry M33388 served as a reference sequence for CYP2D6 and GenBank entry NC_000022, the reference sequence for chromosome 22, served as reference sequences for CYP2D7 and CYP2D8. The latter sequence was used due to sequencing variations in the original reference sequence for CYP2D7 and CYP2D8, M33387. Comparisons were performed using the BLAST 2 SEQUENCES program (bl2seq; http://www.ncbi.nlm.nih.gov/BLAST/bl2seq/wblast2/cgi) and CLUSTALW (http://www.ebi.ac.uk/Tools/clustalw/index/html) at default settings.

**CYP2D6 allelic nomenclature**
Sequences were compared with the *CYP2D6* alleles listed by the *CYP2D6* Nomenclature Committee (http://www.cypalleles.ki.se/CYP2D6.htm, accessed 8/29/2011). The Home Page of the Human Cytochrome P450 (CYP) Allele Nomenclature Committee has reclassified *CYP2D7-2D6* hybrid genes that contain a *CYP2D7*-derived exon 1 (see http://www.cypalleles.ki.se/cyp2d6.htm for details) and this manuscript uses this new classification for the first time. For this reason, all *CYP2D7-2D6* and *CYP2D7-2D6 + CYP2D6* alleles are designated *CYP2D6*\(^*\)13A1-\(^*\)13H according to this new nomenclature. The suffixes designate the switch location of *CYP2D7* to *CYPD6* sequence. Examples of alleles discussed in this article include, *CYP2D6*\(^*\)13A1 and *CYP2D6*\(^*\)13A2 (formerly *CYP2D6*\(^*\)13 and *CYP2D6*\(^*\)77, respectively) which switch from *CYP2D7* to *CYP2D6* in intron 1. *CYP2D6*\(^*\)13B (formerly *CYP2D6*\(^*\)79) switches in exon 2, *CYP2D8*\(^*\)13D (formerly *CYP2D6*\(^*\)78) switches in intron 4, *CYP2D6*\(^*\)13F (formerly *CYP2D6*\(^*\)16 and *CYP2D6*\(^*\)66) switches in intron 7 to intron 8, and *CYP2D6*\(^*\)13H (formerly *CYP2D6*\(^*\)76) switches to *CYP2D6* in exon 9.

**Phenotype prediction**

Phenotype prediction was based on the genotype and the activity listed on the *CYP2D6* Allele Nomenclature Webpage (http://www.cypalleles.ki.se/cyp2d6.htm) and the literature cited below. Phenotype prediction was binned into four categories in a manner similar to that reviewed by Ingelman-Sundberg (Ingelman-Sundberg, 2005) and Kirchheiner et al. (Kirchheiner et al., 2004). It should be noted that other methodologies for phenotype prediction have been described and this is a controversial area of pharmacogenomics (Steimer et al., 2004; Gaedigk et al., 2008). In particular, there is controversy about where to classify the *CYP2D6*\(^*\)2A vs. other *CYP2D6*\(^*\)2 alleles. There is evidence that the *CYP2D6*\(^*\)2 alleles (except *CYP2D6*\(^*\)2A) have reduced function, although this is somewhat substrate dependent (Raimundi et al., 2000; Bapiro et al.,
polymorphism found in \textit{CYP2D6*2A} increases protein production, possibly through increased induction, which compensates for the reduced function caused by the other polymorphisms found in the \textit{CYP2D6*2} alleles, resulting in a function similar to and possibly greater than \textit{CYP2D6*1} (Lovlie et al., 2001; Zanger et al., 2001).

The classification used here is as follows: an ultrarapid metabolizer (UM) had more than two normally functioning alleles (\textit{CYP2D6*1} or \textit{CYP2D6*2A}). An extensive metabolizer (EM) had two normally functioning alleles or one normally functioning allele and 2 reduced function alleles (\textit{e.g. CYP2D6*2, CYP2D6*10, CYP2D6*17, CYP2D6*41}) or two normally functioning alleles and a reduced function allele. An intermediate metabolizer (IM) had one normally functioning allele and either a reduced function allele or a null allele (\textit{e.g. CYP2D6*4, CYP2D6*5, CYP2D6*6}, or single \textit{CYP2D7-2D6} gene). Samples with two or three reduced function alleles were also considered intermediate metabolizers. A poor metabolizer (PM) had only null alleles or a null allele plus a reduced function allele.

\textbf{Results}

\textbf{Summary of Hybrid Genes Found in \textit{CYP2D6*4, CYP2D6*10, Duplication and Homozygous Samples (Table 2)}}

Overall, 8.5\% of the 1390 samples studied had a \textit{CYP2D6*4N + *4-like} or \textit{CYP2D6*68 + *4-like}, 3.2\% had a \textit{CYP2D6*36 + *10}, 0.3\% had a single \textit{CYP2D7-2D6} gene, and 1.8\% had a \textit{CYP2D7-2D6 + CYP2D6, CYP2D7-CYP2D6 + CYP2D6 X N, or CYP2D7-2D6 + CYP2D6-2D7 X 2 + CYP2D6} hybrid tandem arrangement (described below). However, this set of samples was enriched for \textit{CYP2D6*4, CYP2D6*10}, homozygous samples and samples with duplication.
signals. Therefore, the overall frequency is expected to be different in the general population which should contain a random variety of alleles.

Of samples with either a heterozygous or homozygous CYP2D6*4 allele (N=508), 9 (1.8%) had CYP2D6*4N + *4-like and 109 (21.5%) had CYP2D6*68 + *4-like. Of samples with a heterozygous or homozygous CYP2D6*10 allele (N=209), 44 (21.1%) had CYP2D6*36 + *10 (Figure 2F). For homozygous CYP2D6*4 and CYP2D6*10 samples, we did not determine if one or both chromosomes contained a hybrid tandem arrangement. To determine the risk that an individual CYP2D6*4 or CYP2D6*10 allele is actually in a hybrid tandem arrangement, we also calculated the percentage of heterozygous samples which had a hybrid tandem. 7 (1.5%) of CYP2D6*4 heterozygotes (N=482) harbored CYP2D6*4N + *4-like and 83 (17.2%) harbored CYP2D6*68 + *4-like. 28 (14.5%) of CYP2D6*10 heterozygous samples (N=193) harbored CYP2D6*36 + *10.

Of samples homozygous for any allele (N=332), 4 (1.2%) had a CYP2D7-2D6 gene. Of samples with a duplication signal (N=341), 25 (7.3%) contained a CYP2D7-2D6 + CYP2D6 hybrid tandem or similar arrangement as discussed below.

**Characteristics of 5’ Genes in CYP2D7-2D6 + CYP2D6 Hybrid Tandem**

Table 3 shows the actual 5’ genes present in the CYP2D7-2D6 + CYP2D6 hybrid tandems we observed. DNA sequence analysis was done for all of the 5’ genes found in this research. In 7 samples the DNA sequence analysis revealed that the 5’ gene was actually a CYP2D7 which converted to a CYP2D6 sequence in the 0.6kb repeat located just downstream of the CYP2D7 sequence and upstream of the rep dup sequence (Figures 1A and 2C). Four 2D7 + CYP2D6*I (cases 1, 10, 20, and 22), two 2D7 + *4I (case 4 and 16), and one 2D7 + CYP2D6*4 (case 8) were observed.
Twelve CYP2D6*13A2 + *2A, one CYP2D6*13A1 (variant 1) + *1 (case 7), one CYP2D6*13A1 (variant 2) + CYP2D6*1 X N (case 15), two CYP2D6*13D + *2A (cases 13 and 18), one CYP2D6*13G2 + *1 (case 5), and one CYP2D6*13B + *68 X 2 + *4 (case 19) were also observed. GenBank entry numbers for CYP2D6*13A1 variant 1 and variant 2 are HQ670230 and HQ670231, respectively. CYP2D6*13A1 variant 1 has intron 4 c.667-149G while CYP2D6*13A1 has a T and variant 1 has an intron 5 c.844-49G while CYP2D6*13A1 has a T. CYP2D6*13A1 variant 2 has only the intron 4 c.667-149G. (Note that locations are derived from the CYP2D6 reference sequence M33388.) An additional novel allele was identified which was assigned CYP2D6*13G2 by the nomenclature committee, GenBank entry number HQ670229). In this sample, the conversion from CYP2D7 to CYP2D6 occurred in intron 7 based upon sequence data and comparison to CYP2D7 and CYP2D6 sequence data although it also contains one CYP2D7-like SNP in exon 8.

To confirm the presence of a rep dup in each of these samples, DNA sequence analysis was performed using amplicon L (Table 1, Figure 1B and 2B-D). In every case, the first 53 nucleotides of rep dup contained the five nucleotides specific to rep 6 and the last 39 nucleotides of rep dup contained the four nucleotides that define rep 7 (Steen et al., 1995; Soyama et al., 2006). Absence of the CYP2D7 pseudogene 5’ to all CYP2D7-2D6 hybrids was confirmed by the presence and size of fragment J (Figure 2B-2E) which cannot be generated if CYP2D7 is present due to the size of the amplicon.

Table 4 summarizes the frequency in which a given duplicated allele was found to have an associated CYP2D7-2D6 gene. A CYP2D7-2D6 hybrid tandem was present in 7 of 91 (7.7%) samples involving a CYP2D6*1 allele. Similarly, a hybrid tandem was present in 14 of 98 (14.3%) of samples involving a CYP2D6*2A, 2 of 25 (8%) alleles involving a CYP2D6*4 allele,
and 2 of 4 (50%) samples involving a CYP2D6*41 allele. Notably, there were 20 samples where a CYP2D6*2 X N (not CYP2D6*2A X N) was predicted to be duplicated and none of them were found to have a tandem hybrid. While these numbers are small, the risk of an undetected hybrid tandem being present in a sample appears to be greatest in those samples where the CYP2D6*1, CYP2D6*2A, CYP2D6*4 and CYP2D6*41 is predicted to be duplicated using the CYP2D6 ASPE kits as described here.

Characteristics of Single CYP2D7-2D6 Genes

DNA sequencing was performed on all homozygous samples which were found to contain a single CYP2D7-2D6 gene (Figure 2A). Of these, 3 were found to be a CYP2D6*13F and one was found to have a novel allele named CYP2D6*13G1 (JN618990).

Gene Copy Numbers of CYP2D7-2D6 + CYP2D6 Hybrid Tandems (Table 3)

The CYP2D6 copy number was determined for all samples that gave a duplication signal and were determined to have a CYP2D7-2D6 + CYP2D6 arrangement. Real time PCR (rtPCR) assays were performed which used one primer pair plus a Taqman® probe to target the intron 6 region of CYP2D6 and another primer pair plus Taqman® probe to target the 5’ flanking region of the CYP2D6 promoter region (Figure 1 and 2). CYP2D7-2D6 genes were detected on the basis that the copy number varied between the two regions. Specifically, CYP2D7-2D6 genes have a CYP2D7 promoter region so the 5’ flanking probe will not bind there in distinction to any CYP2D6 genes present in a sample. CYP2D7-2D6 genes that switch to CYP2D6 sequence upstream of intron 6 have a binding site for the intron 6 probe. Therefore, early crossing CYP2D7-2D6 genes (e.g. CYP2D6*13A1, CYP2D6*13A1 variant 1 and variant 2, CYP2D6*13A2, CYP2D6*13D, CYP2D6*13B) which convert to CYP2D6 sequence upstream of intron 6 were detected with the intron 6 probe but not by the 5’ CYP2D6 probe. The CYP2D7-
2D6 gene which was not detected by the intron 6 probe was CYP2D6*13G2 due to its conversion to CYP2D6 in intron 7. Neither probe detected instances where the 5’ CYP2D7 pseudogene converted to the CYP2D6 trailing sequence in the 0.6kb repeat region just downstream of CYP2D7 and upstream of the rep dup sequence.

CYP2D6-2D7 hybrids (cases 19 and 24) were detected with the CYP2D6 5’ probe because they have a CYP2D6 promoter sequence (Figure 2F). The intron 6 assay did not yield an amplification product from the CYP2D6*68 hybrid because this allele switched to CYP2D7 sequence in intron 1. Copy numbers for each assay were as expected for all but 2 samples (Table 3). Five - six copies of CYP2D6 were observed with both the intron 6 and the CYP2D6 promotor probes for case 15 indicating that CYP2D6*1 is multiplied but the total number could not be determined. The product literature states that decreased confidence estimates are observed with increased copy numbers. Also, case 19 had 3 and 4 copies with the intron 6 and CYP2D6 promotor assay, respectively, suggesting 2 copies of CYP2D6*68. Cases 15 and 19 are discussed in detail below.

**Novel CYP2D7-2D6 + CYP2D6 X N Hybrid Tandem**

We could not reliably determine the number of CYP2D6 genes by rtPCR in case 15 (Table 3). We expected that the intron 6 assay would show one more copy than the CYP2D6 promotor assay, since the CYP2D6*13A1 gene would not be detected by the Taqman® probe in the promotor region. The presumed reason for this finding is that reliability of rtPCR to detect copy numbers higher than 5 is poor (see product literature). The structure of this sample was supported by CYP2D6 ASPE kit analysis of fragments A and B (Table 1, Figure 2D), both of which were genotyped as CYP2D6*1. This is the first time that a CYP2D7-2D6 hybrid gene has been reported as part of a multiplied CYP2D6 arrangement.
Novel CYP2D7-2D6 + CYP2D6-2D7 X 2 + CYP2D6 Hybrid Tandem

The results from PCR fragment analysis (Figure 3), CYP2D6 ASPE kit, sequence analysis (amplicons E, G and K), and copy number assays were compiled to generate the structure of case 19 (Table 3, Figure 2E), a novel CYP2D7-2D6 + CYP2D6-2D7 X 2 + CYP2D6 hybrid tandem.

The presence of CYP2D6*13B followed by CYP2D6*68 on the same chromosome was confirmed by the amplification of fragment K (Figure 3, lane 7) which spans from exon 9 in CYP2D6 to exon 9 in CYP2D7. This amplicon was sequenced using multiple amplicons and primers to ensure continuity between the CYP2D6*13B and CYP2D6*68 alleles. In addition, four copies of the CYP2D6 promoter region were observed with the rtPCR CYP2D6 promoter copy number assay, consistent with two CYP2D6*68 alleles plus CYP2D6*4 on one chromosome and CYP2D6*1 on the other chromosome. Whereas, 3 copies of CYP2D6 intron 6 were found using the intron 6 copy number assay. This is because CYP2D6*13B contains CYP2D6 intron 6 as does CYP2D6*4 on the same chromosome and CYP2D6*1 on the other chromosome but CYP2D6*68, which converts to CYP2D7 in intron 1, is not detected by the intron 6 copy number assay. The presence of tandem CYP2D6*68 alleles was confirmed by PCR amplification using forward primers located in the rep 7 (fragment M) and rep 6 (fragment N) regions of the gene with a reverse primer located in exon 9 of CYP2D7 (Figure 3, lane 10). These amplifications would occur only if both rep dup and rep 7 each were followed by a CYP2D6*68 (CYP2D6-2D7) arrangement (see Figure 2E). Original genotyping of case number 19 yielded a CYP2D6*1/*4 genotype. This genotype is consistent with the arrangement shown in Figure 2E because the locations of the Luminex kit primers allows PCR amplification of both the normal CYP2D6 (Figure 1A) as well as the 3’ most CYP2D6 structure in the novel arrangement (Figure 2E). It should be noted that although CYP2D6*13B has been observed
alone (Gaedigk et al., 2010b), this is the first time it has been reported in tandem with other genes.

**Impact of Hybrid Genes on Predicted Phenotype**

*CYP2D6-CYP2D7 + CYP2D6* Hybrid Tandems

The *CYP2D6*4, *CYP2D6*4N, and *CYP2D6*4-like alleles have the 1846A variation which results in a splicing defect which encodes an enzyme with no activity. The *CYP2D6*68 converts to *CYP2D7* sequence from intron 1 forward and is also expected to produce a null allele due to the presence of the *CYP2D6*10 polymorphism c.100C>T plus the presence of an exon 4 c.631dupG resulting in a frame shift causing a stop codon 43 bases downstream on the insertion (p.Glu211GlyfsX43) although no *in vivo* or *in vitro* research has been reported for this allele. The *CYP2D6*10 allele is described as having reduced function and the *CYP2D6*36 allele is described as having poor function (Gaedigk et al., 2006). Thus, the presence of a *CYP2D6*36, *4N*, and *68* in a tandem does not impact the predicted phenotype of the sample.

Single *CYP2D7-2D6* Samples

All *CYP2D7-2D6* genes carry a T-insertion in exon 1 that causes a frameshift and a premature termination (c.137_138insT, p.Leu47AlafsX207). These hybrids are nonfunctional (Gaedigk et al., 2010b). Four (100%) of 4 samples originally genotyped as having a homozygous genotype which were found to have a *CYP2D7-2D6* gene on one chromosome (Figure 2A) had a change in phenotype prediction. These samples were originally genotyped as *CYP2D6*1/*1 but were found to actually be heterozygous for *CYP2D6*13G1 (N=1) and *CYP2D6*13F (N=3) resulting in a change in phenotype prediction from EM to IM.

*CYP2D7-2D6 + CYP2D6* Hybrid Tandem
Table 3 shows the change in phenotype of the samples with *CYP2D7-2D6* + *CYP2D6* hybrid tandems (Figure 2B and 2C) observed in this research. Detection of the hybrid gene resulted in a change of predicted phenotype in 23 (92%) of 25 samples containing a hybrid. In 9 of the samples (cases 4, 11, 13, 14, 19, 20, 22, 24 and 25), the duplicated allele could not be predicted at the time of original genotyping and the original phenotype was specified “EM or IM”. After the correct genotype was determined, the predicted phenotype was confirmed to be IM. Thus, in these cases, more precise phenotype prediction could be determined using the methods described here. If these 5 cases are also included in the “no change” group, then 18 samples (56%) had a change in phenotype.

**Discussion**

The intent of this study was to determine the frequency of *CYP2D6-2D7* and *CYP2D7-2D6* genes in samples containing *CYP2D6*/*4* alleles, *CYP2D6*/*10* alleles, homozygous samples of any variety and samples with duplication signals that were originally genotyped with a commercially available ASPE kit. Furthermore, we wanted to determine the impact of these undetected genes on predicted phenotype.

The main findings of this research are that hybrid genes: 1) are relatively common in clinical samples containing *CYP2D6*/*4* alleles, *CYP2D6*/*10* alleles, and duplication signals, 2) are uncommon in homozygous samples, and 3) frequently impact phenotype prediction in samples originally genotyped as homozygous and, separately, those that were originally genotyped as having a duplication. However, hybrid tandems do not change predicted phenotype in *CYP2D6*/*4* and *CYP2D6*/*10*-containing samples.

Our results involving duplication samples supports those described by Gaedigk et al (Gaedigk et al., 2010a) in which three *CYP2D6*/*13A2* + *2* and one *CYP2D6*/*13D* + *2* were found in 32
duplication-positive Caucasian samples (12.5%) and one CYP2D6*13H + *I was found in 59
duplication-positive African-American samples (1.6%) for an overall frequency of 5.4%
compared to our 7.3%. The difference between these two studies is likely due to difference in
the ethnic mix of patients, although ethnic and racial data is not available for our samples. We
also saw more CYP2D6*13A2 + *2A (N=12) vs. CYP2D6*13D + *2A (N=2) (Table 3).
Interestingly, no hybrid genes were observed with other CYP2D6*2 variants. We also identified
seven samples with duplication signals which contained a CYP2D7 gene with a CYP2D6-like
region downstream of exon 9 similar to that described by Gaedigk et al (Gaedigk et al., 2010b).
Specifically, cases 1, 10, 20, and 22 had a CYP2D7 + *I, cases 4 and 16 had a CYP2D7 + *4I,
and case 8 had a CYP2D7 + *4 and all had sequence confirmed rep dup 3’ to the CYP2D7 hybrid
allele.

In our collection of 341 samples originally genotyped with duplications, a hybrid tandem gene
was present in 25 (7.3%) and this had an impact upon phenotype prediction on 92% of these
samples. Furthermore, the presence of a single CYP2D7-2D6 gene was only found in 4 of 332
samples originally genotyped as homozygous but it changed predicted phenotype in 100% of
cases. Although the need to test for the rare single CYP2D7-2D6 gene in a clinical setting can be
debated in the absence of actual evidence that a given patient has genotype-phenotype
discordance, the authors argue that the frequency of hybrid tandems in duplication samples and
their impact on phenotype prediction warrants their analysis.

The CYP2D6 ASPE kit that was used to originally genotype our samples was not designed to
detect hybrid genes thus it is not surprising that none were detected by this kit during original
clinical genotyping. Hybrid genes will not be detected in a sequence-based analysis either unless
primers are designed to specifically allow for the PCR amplification of these hybrids.
Furthermore, microarray-based assays, such as the AmpliChip CYP450 Test (Roche, Basel, Switzerland) will not detect hybrids genes unless the initial PCR amplification step uses primers that amplify the hybrid gene and then hybrid gene-specific probes are designed into the array.

Another feature that can impact phenotype is gene multiplication. In this study, 2/25 (8%) of the samples tested had otherwise undetected multiplications. This allowed for more accurate phenotype prediction in case 15 ($CYP2D6^{*1/*13A1 \text{ var 2}} + ^{*1 X N}$) and the rtPCR copy number assays allowed for the identification of two unique $CYP2D7-2D6$ hybrid arrangements (case 15 and 19). These findings show that duplication samples are challenging to accurately genotype and it appears that using an array of analyses, such as we have done here, is necessary to be sure of completely accurate genotyping.

Limitations of this study include: No ethnic or racial data was available for these sample because, per IRB requirement for this type of study, all samples needed to be de-identified. Similarly, no phenotypic data is available for these individual samples.

Conclusions

$CYP2D7-2D6$ genes must be specifically tested for using methods comparable to those outlined here in order to be detected. $CYP2D7-2D6$ genes were found frequently in samples originally genotyped to have duplications using a $CYP2D6$ ASPE kit. Their detection resulted in a change in phenotype prediction in a high percentage (92%) of cases where a duplication call was made. Furthermore, phenotype changed in 100% of rarer cases where a single $CYP2D7-2D6$ gene was present in our samples.

$CYP2D6-2D7$ genes must also be specifically tested for to be detected but they have no effect upon phenotype prediction because they were found in tandem with $CYP2D6^{*4}$ and $CYP2D6^{*10}$.
alleles which are null or have reduced function, respectively, and the hybrid genes (CYP2D6*36, CYP2D6*4N, and CYP2D6*68) have poor phenotype activity.

Two new structures, CYP2D6*13B + (*68 X 2) + *4 and CYP2D6*13A1 variant 2 + (*1 X N) were described as were two new alleles, CYP2D6*13G1 and CYP2D6*13G2, and two variations of the CYP2D6*13A1 allele.

The information presented here is crucial for detection of hybrid genes regardless of the platform used to genotype CYP2D6. Probe sets for microarray detection must be designed with knowledge of hybrid genes or probes will not bind in a fashion that will allow their detection. In addition, CYP2D6 full gene sequencing and high throughput genotyping using next generation DNA analyzers will need to use the DNA sequences and genomic structures presented here as scaffolds for interpretation of sequence results.
Authorship Contributions:

Participated in research design: Black, Walker, O’Kane

Conducted experiments: Walker, Harmandayan

Contributed new reagents or analytic tools: Black, Walker, O’Kane

Performed data analysis: Black, Walker, O’Kane, Harmandayan

Wrote or contributed to writing of the manuscript: Black, Walker, O’Kane, Harmandayan
References:


Gaedigk A, DiAnne Bradford L, Alander SW and Steven Leeder J (2006) CYP2D6*36 gene arrangements within the CYP2D6 locus: Association of CYP2D6*36 with poor metabolizer status. *Drug Metabolism and Disposition* **34:**563-569.


Footnotes

Financial Support: This research was supported by National Institutes of Health/National Center for Research Resources Clinical and Translational Science Awards [Grant Number UL1 RR024150]. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Health. Funding for this research was also provided by the Mayo Clinic.

Financial Disclosure: Drs. Black and O’Kane and the Mayo Clinic have licensed intellectual property to AssureRx.
Legends for Figures:

Figure 1: *CYP2D* locus structures for single, typical duplicated and deleted arrangements

(A) Single *CYP2D6* arrangement. *CYP2D8* (not shown) and *CYP2D7* pseudogenes are located 5’ to the *CYP2D6* gene. Similar 0.6kb repeats follow the *CYP2D7* and *CYP2D6* sequences as do rep 7 and rep 6, respectively, which differ by just a few nucleotides in the rep’s 5’ and 3’ regions. Note that the 1.6kb spacer is located 3’ to the *CYP2D7* pseudogene which is absent downstream of the *CYP2D6* gene. (B) Typical *CYP2D6* duplication arrangement. The first *CYP2D6* gene is followed by rep dup, a hybrid containing a 5’ rep 6 sequence and a 3’ rep 7 sequence. Multiplications of the sequence shown between the brackets are known to exist. (C) *CYP2D6* deletion arrangement (*CYP2D6*<sup>*6</sup>*5) where *CYP2D7* is followed by a rep del which is a hybrid containing a 5’ rep 7 sequence and a 3’ rep 6 sequence. PCR fragments used in this investigation are depicted as lettered lines under the structures (see Table 1). Probe locations for Taqman® copy number assays are designated as * and + for the 5’ flanking *CYP2D6* assay and the *CYP2D6* intron 6 assay, respectively.

Figure 2: *CYP2D* locus structures for single and hybrid tandems of *CYP2D7-2D6* and *CYP2D6-2D7*.

(A) Single *CYP2D7-2D6* arrangement (e.g. *CYP2D6*<sup>*13</sup>A1, *CYP2D6*<sup>*13</sup>F, *CYP2D6*<sup>*13G1</sup>). (B) *CYP2D7-2D6* + *CYP2D6* arrangement (e.g. *CYP2D6*<sup>*13</sup>A1, *CYP2D6*<sup>*13</sup>A2, *CYP2D6*<sup>*13D</sup>, *CYP2D6*<sup>*13G2</sup> with a tandem *CYP2D6*<sup>*1</sup>, *CYP2D6*<sup>*2A</sup>). Notice that the hybrid gene is followed by rep dup. (C) *CYP2D7* gene in a duplication arrangement that is followed by rep dup upstream of the tandem *CYP2D6* gene (e.g. *CYP2D6*<sup>*1</sup>, *CYP2D6*<sup>*4</sup>, *CYP2D6*<sup>*41</sup>). (D) Novel *CYP2D7-2D6* + *CYP2D6 X N* hybrid tandem multiplication arrangement. In this case the *CYP2D7-2D6* hybrid is a *CYP2D6*<sup>*13</sup>A1 variant 2 and the *CYP2D6* alleles are *CYP2D6*<sup>*1</sup>. The
absolute number of CYP2D6*1 alleles in tandem could not be determined. (E) Novel CYP2D7-2D6 hybrid (CYP2D6*13B) followed by rep dup, CYP2D6-2D7 hybrids (CYP2D6*68 X 2) and a CYP2D6 gene (CYP2D6*4-like). The ←M and M→ designations show the relative locations of fragment M primers in this arrangement, which generates a product because the CYP2D6-2D7 hybrid is found in tandem in this arrangement. (F) CYP2D6-2D7 + CYP2D6 arrangement (e.g. CYP2D6*36, CYP2D6*4N, CYP2D6*68 with a tandem. CYP2D6*10, CYP2D6*4-like). Note that the CYP2D6-2D7 gene is followed by rep 7 rather than rep dup. Multiduplications have multiples of the sequence shown between the brackets and were observed in (D) and (E). PCR fragments used in this analysis are depicted as lettered lines under the structures (Table 1). Probe locations for Taqman® copy number assays are designated as * and + for the 5’ flanking CYP2D6 assay and the CYP2D6 intron 6 assay, respectively. Note that CYP2D8 is thought to be present in all of the structures but is only shown in structures Figure 2A-C for simplicity.

Figure 3: Amplicon analysis of novel CYP2D6*13B + *68 X 2 + *4

PCR fragment analysis for case number 19 (Figure 2E). PCR primers and expected fragment sizes for products D, E, G, J, K and M are listed in Table 1. Fragment analysis was performed using the Agilent Technologies DNA 12000 kit. The ladder in lane 1 shows the size in base pairs of the molecular weight marker. For each amplicon, a negative control is shown first followed by amplicons generated for case number 19. Amplicon J (Lane 2 and 3) primed from the trailing sequence of CYP2D8 to the beginning of rep 6 or rep dup. When the CYP2D7 pseudogene was present, no amplicon was generated due to the lengthy span of DNA. In this case (Lane 3), a fragment was generated indicating that CYP2D7 was absent. Amplicon G (Lane 4 and 5) uniquely amplified the CYP2D7-2D6 gene (CYP2D6*13B) and was used in sequencing. Amplicon K (Lane 6 and 7) primed from the CYP2D6 exon 9 sequence in CYP2D6*13B to the
CYP2D7 exon 9 sequence in CYP2D6*68. This amplicon was sequenced using multiple amplicons to ensure continuity between the CYP2D6*13B and CYP2D6*68 alleles. Amplicon M (Lane 8 and 9) primes from a unique forward region in rep 7 to the exon 9 sequence in CYP2D7 found in CYP2D6*68. Amplification would only occur with a complete rep 7 (not rep dup) upstream of CYP2D6*68. Amplicon E (Lane 10 and 11) uniquely amplified CYP2D6-2D7 genes (CYP2D6*68) and was used to generate an amplicon which was genotyped using the CYP2D6 v2 kit thus verifying the CYP2D6*68 allele. Amplicon D (Lane 12 and 13) amplified CYP2D6 gene in any normal arrangement (e.g. Figure 1A), any CYP2D6 gene 3’ allele in a duplication arrangement (Figure 1B), and any CYP2D6 gene in a tandem (Figure 2B – 2F). This amplicon generated the genotype of heterozygous CYP2D6*4 from the tandem gene in Figure 2E and the CYP2D6*1 allele on the other chromosome using the CYP2D6 v2 kit.
Table 1: PCR Primers and Amplicon Characteristics

<table>
<thead>
<tr>
<th>PCR product</th>
<th>Sequence</th>
<th>Arrangement Amplified</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>GTCCCACACCAGGCACCTGTACT</td>
<td>Nonduplicated gene &amp; deletion arrangement</td>
<td>15629, 3471</td>
</tr>
<tr>
<td></td>
<td>GAATTAGTGGTGTTGGGTGTTTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>TCACCCCCAGGGAGCTATATCA</td>
<td>5' gene of duplication arrangement</td>
<td>12052</td>
</tr>
<tr>
<td></td>
<td>CCACAGCCTCAATAAGTGAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>CACCCACCTACGGCTCGTC</td>
<td>3' gene of duplication arrangement</td>
<td>12103</td>
</tr>
<tr>
<td></td>
<td>TAGGTAGGCTCCTGCTTAGCTCCCTGACGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>TGCCACATTATCCGCCTGGAATA</td>
<td>Any normal arrangement of CYP2D6</td>
<td>8433</td>
</tr>
<tr>
<td></td>
<td>TAGGTAGCCCTGCGCTATAGCTCCCTGACGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>TACCCACCCAGGGAGCTATATCA</td>
<td>CYP2D6-2D7 gene</td>
<td>6714</td>
</tr>
<tr>
<td></td>
<td>TACGGTGGCCTCCCTGCGAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>CCTGTGTTGGGCTTGGGAGCCTTG</td>
<td>CYP2D7-2D6 gene</td>
<td>5742</td>
</tr>
<tr>
<td></td>
<td>TGTGTGAGGAGGCTGCCTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>ACCTGGACCGCTACTTTTA</td>
<td>PCR fragment from CYP2D8-unique 3' region to CYP2D6 3' region</td>
<td>9525</td>
</tr>
<tr>
<td></td>
<td>TAGGTAGCCCTGCGCTATAGCTCCCTGACGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>GCCACCATGCTTGGTCTTGTGCCTTC</td>
<td>PCR fragment from exon 9 in CYP2D6 to exon 9 in CYP2D7.</td>
<td>12152</td>
</tr>
<tr>
<td></td>
<td>AAAGCTGACGACGAGAGTGGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>TCAGCCTGTCACCTCACCACAGG</td>
<td>Any rep dup region</td>
<td>5410</td>
</tr>
<tr>
<td>---</td>
<td>------------------------</td>
<td>-------------------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>CCACAGCCCTCAATAAGTGAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>CCTCAGGGATGCTGCTGTCTG</td>
<td>PCR fragment from a unique region in rep 7 to a CYP2D7 specific exon 9 sequences and only amplifies if there are two rep 7 regions.</td>
<td>11432</td>
</tr>
<tr>
<td></td>
<td>AAAGCTGACGACACGAGAGTGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>CCTCAGGGACGCTGCTGTACA</td>
<td>PCR fragment from a unique region in rep 6 to a CYP2D7 specific exon 9 sequence</td>
<td>11432</td>
</tr>
<tr>
<td></td>
<td>AAAGCTGACGACGAGAGTGCTGA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1: PCR primers and amplicon characteristics**

The primers shown were used to generate PCR amplicons for testing by allele-specific primer extension assay, DNA sequencing or for PCR fragment analysis. Refer to Figures 1 and 2 for approximate annealing sites of these primers on the various genomic structures.
## Table 2: Frequency of Hybrid Genes Per Sample Type

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterozygous or homozygous CYP2D6*4</td>
<td>508</td>
<td></td>
</tr>
<tr>
<td>Containing CYP2D6*4N + *4-like</td>
<td>9</td>
<td>1.8%</td>
</tr>
<tr>
<td>Containing CYP2D6*68 + *4-like</td>
<td>109</td>
<td>21.5%</td>
</tr>
<tr>
<td>Percentage of total samples (N=1390) with</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2D6*4N + <em>4-like or CYP2D6</em>68 + *4-like</td>
<td></td>
<td>8.5%</td>
</tr>
<tr>
<td>Heterozygous CYP2D6*4</td>
<td>482</td>
<td></td>
</tr>
<tr>
<td>Containing CYP2D6*4N + *4-like</td>
<td>7</td>
<td>1.5%</td>
</tr>
<tr>
<td>Containing CYP2D6*68 + *4-like</td>
<td>83</td>
<td>17.2%</td>
</tr>
<tr>
<td>Heterozygous CYP2D6*10</td>
<td>209</td>
<td></td>
</tr>
<tr>
<td>Containing CYP2D6*36 + *10</td>
<td>44</td>
<td>21.1%</td>
</tr>
<tr>
<td>Percentage of total samples (N=1390) with</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2D6*36 + *4-like</td>
<td></td>
<td>3.2%</td>
</tr>
<tr>
<td>Heterozygous CYP2D6*10</td>
<td>193</td>
<td></td>
</tr>
<tr>
<td>Containing CYP2D6*36 + *10</td>
<td>28</td>
<td>14.5%</td>
</tr>
<tr>
<td>Any homozygous genotype</td>
<td>332</td>
<td></td>
</tr>
<tr>
<td>Containing CYP2D7-2D6 hybrid genes</td>
<td>4</td>
<td>1.2%</td>
</tr>
<tr>
<td>Percentage of total samples (N=1390) with</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2D7-2D6 hybrid gene</td>
<td></td>
<td>0.3%</td>
</tr>
</tbody>
</table>
Any duplication sample | 341 |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Containing CYP2D7-2D6 + CYP2D6 hybrid tandem present</td>
<td>25</td>
</tr>
<tr>
<td>Percentage of total samples (N=1390) with CYP2D7-2D6 + CYP2D6 or novel structure (N=25)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Frequency of hybrid genes per sample type

The frequency of hybrids genes in CYP2D6*4 and CYP2D6*10 samples is reported for both heterozygous or homozygous and heterozygous alone. We did not determine whether one or both chromosomes contained a hybrid tandem in homozygous samples. For this reason, to determine the risk that a CYP2D6*4 or CYP2D6*10 allele is actually in a hybrid tandem arrangement, we determined the percentage of heterozygous samples for each which had a hybrid tandem. The “Any Homozygous Genotype” column included any sample with an initial homozygous genotype except those with duplications. Finally, the frequency of hybrid genes in samples originally genotyped as having duplications is shown. The percentage of hybrid alleles in the 1390 samples is also shown by category although note that the sample set was not from a random population but was purposefully enriched for samples with CYP2D6*4, CYP2D6*10, duplication and homozygous genotypes thus frequency of hybrid genes in a general populations is likely to be different.
Table 3: Characteristics of Duplication Samples Found to Contain CYP2D7-2D6 Genes

<table>
<thead>
<tr>
<th>Case Number</th>
<th>Original CYP2D6 Genotype</th>
<th>Copy Number per Taqman® Results</th>
<th>Predicted Phenotype Comparing Original Genotype to Actual Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CYP2D6 Promoter Assay</td>
<td>CYP2D6 Intron 6 Assay</td>
</tr>
<tr>
<td>1</td>
<td>*1/*1 X N</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>*2A X N/*17</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>*1/*2A X N</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>(*2A/*41) X N#</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>*1 X N/*2A</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>*2A/*2A X N</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>*1 X N/*5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>*2A X N/*4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>*2A/*2A X N</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>*1 X N/*2A</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>*2A/*41 X N#</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>*2A X N/*4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Characteristic</td>
<td>Genotype</td>
<td>Impact</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------</td>
<td>----------</td>
<td>--------------</td>
</tr>
<tr>
<td>13</td>
<td>(*2A/*41) X N#</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>(*2A/*4) X N</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>*1/*1 X N</td>
<td>5-6</td>
<td>5-6</td>
</tr>
<tr>
<td>16</td>
<td>*41/*41 X N</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>*2A/*2A X N</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>(*1/*2A) X N</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>(*1/*4) X N</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>(*1/*41) X N#</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>(*1/*2A) X N</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>22</td>
<td>(*1/*6) X N</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>*2A/*2A X N</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>24</td>
<td>(*2A/*4) X N</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>(*2A/*41) X N</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 3: Characteristics of duplication samples found to contain CYP2D7-2D6 genes**

In the original CYP2D6 genotype column, parentheses around the star alleles designate those samples in which the duplicated allele could not be determined in the clinical assay. Correspondingly, this led to some ambiguity in the original phenotype prediction for a subset of samples. Samples originally genotyped as a CYP2D6*2 which were later found to be CYP2D6*41 are designated with a (#). This occurred due to limitations of the CYP2D6 kit v1. Subsequent testing was done by CYP2D6 kit v2 which does detect
CYP2D6*41. Taqman® copy number assay probe locations for these structures are found in Figure 1 and 2. Confidence estimates for the copy number assays were all >0.99 with the following exceptions: 1^0.56, 2^<0.50, 3^0.85.
Table 4: Frequency of CYP2D7-2D6 Genes in Tandem to Specific CYP2D6 Alleles

<table>
<thead>
<tr>
<th>Predicted duplicated CYP2D6 allele</th>
<th>N</th>
<th>CYP2D7-2D6 gene present</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>*1</td>
<td>91</td>
<td>7</td>
<td>7.7</td>
</tr>
<tr>
<td>*2</td>
<td>20</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>*2A</td>
<td>98</td>
<td>14</td>
<td>14.3</td>
</tr>
<tr>
<td>*4</td>
<td>25</td>
<td>2</td>
<td>8.0</td>
</tr>
<tr>
<td>*9</td>
<td>1</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>*10</td>
<td>1</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>*41</td>
<td>4</td>
<td>2</td>
<td>50.0</td>
</tr>
<tr>
<td>Total</td>
<td>240</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Frequency of CYP2D7-2D6 genes in tandem to specific CYP2D6 alleles

This table shows the frequency in which a predicted duplicated allele was found to have a CYP2D7-2D6 + CYP2D6 arrangement. Star alleles predicted to be duplicated using the CYP2D6 ASPE kits are shown in column 1.
Figure 1

A. CYP2D8

B. CYP2D7

C. CYP2D8

This article has not been copyedited and formatted. The final version may differ from this version.
Figure 2

A. CYP2D8

B. CYP2D8

C. CYP2D8

D. CYP2D8

E. CYP2D8

F. CYP2D8


(*1, *2A)

(*1, *4, *41)

(*1)

(*1)

(*13B)

(*68)

(*4-like)

(*36, *4N, *68)

(*10 and *4-like)
Correction to “Frequency of Undetected CYP2D6 Hybrid Genes in Clinical Samples: Impact on Phenotype Prediction”

In the article referenced above [Black JL 3rd, Walker DL, O’Kane DJ, and Harmandayan M (2012) Drug Metab Dispos 40:111–119], Table 2, column 1, row 11 should read “Percentage of total samples (n = 1390) with CYP2D6*36 + *10-like (n = 44).” The corrected table is shown below.

In addition, the assignments of allele designations by the CYP2D6 Nomenclature Committee to the CYP2D7-2D6 hybrid allelic variants are undergoing revision. Readers are encouraged to refer to http://www.cypalleles.ki.se/cyp2d6.htm for the most recent nomenclature.

The online versions of this article have been corrected in departure from the original version.

The authors regret this error and apologize for any confusion or inconvenience it may have caused.

### TABLE 2
Frequency of hybrid genes per sample type

The frequency of hybrids genes in CYP2D6*4 and CYP2D6*10 samples is reported for both heterozygous or homozygous and heterozygous alone. We did not determine whether one or both chromosomes contained a hybrid tandem in homozygous samples. For this reason, to determine the risk that a CYP2D6*4 or CYP2D6*10 allele is actually in a hybrid tandem arrangement, we determined the percentage of heterozygous samples for each, which had a hybrid tandem. The “Any homozygous genotype” row included any sample with an initial homozygous genotype except those with duplications. Finally, the frequency of hybrid genes in samples originally genotyped as having duplications is shown. The percentage of hybrid alleles in the 1390 samples is also shown by category, although note that the sample set was not from a random population but was purposefully enriched for samples with CYP2D6*4, CYP2D6*10, duplication, and homozygous genotypes; thus, frequency of hybrid genes in a general population is likely to be different.

<table>
<thead>
<tr>
<th>Category</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterozygous or homozygous CYP2D6<em>4 Containing CYP2D6</em>4N + *4-like</td>
<td>508</td>
<td>1.8</td>
</tr>
<tr>
<td>Containing CYP2D6*68 + *4-like</td>
<td>109</td>
<td>21.5</td>
</tr>
<tr>
<td>Percentage of total samples (n = 1390) with CYP2D6*4N + <em>4-like or CYP2D6</em>68 + *4-like (n = 118)</td>
<td></td>
<td>8.5</td>
</tr>
<tr>
<td>Heterozygous CYP2D6<em>4 Containing CYP2D6</em>4N + *4-like</td>
<td>482</td>
<td>1.5</td>
</tr>
<tr>
<td>Containing CYP2D6*68 + *4-like</td>
<td>83</td>
<td>17.2</td>
</tr>
<tr>
<td>Heterozygous or homozygous CYP2D6<em>10 Containing CYP2D6</em>36 + *10</td>
<td>209</td>
<td>21.1</td>
</tr>
<tr>
<td>Percentage of total samples (n = 1390) with CYP2D6*36 + *10-like (n = 44)</td>
<td></td>
<td>3.2</td>
</tr>
<tr>
<td>Heterozygous CYP2D6*10</td>
<td>193</td>
<td>14.5</td>
</tr>
<tr>
<td>Containing CYP2D6*36 + *10</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Any homozygous genotype Containing CYP2D7-2D6 hybrid genes</td>
<td>332</td>
<td>1.2</td>
</tr>
<tr>
<td>Percentage of total samples (n = 1390) with CYP2D7-2D6 hybrid gene (n = 4)</td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>Any duplication sample Containing CYP2D7-2D6 hybrid tandem present</td>
<td>341</td>
<td>7.3</td>
</tr>
<tr>
<td>Percentage of total samples (n = 1390) with CYP2D7-2D6 + CYP2D6 or novel structure (n = 25)</td>
<td>25</td>
<td>1.8</td>
</tr>
</tbody>
</table>