Establishment of CYP2D6 Reference Samples by Multiple Validated Genotyping Platforms

Hua Fang, Ph.D.[1,2], **Xiao Liu, M.S.**[2], **Jacqueline Ramírez, M.S.**[2], **Noura Choudhury, B.S.**[2], **Michiaki Kubo, Ph.D.**[7], **Hae Kyung Im, Ph.D.**[5], **Anuar Konkashbaev, M.S.**[3], **Nancy J. Cox, Ph.D.**[1,3,6], **Mark J. Ratain, M.D.**[1,2,6], **Yusuke Nakamura, M.D., Ph.D.**[1,2,4,6], and **Peter H. O'Donnell, M.D.**[1,2,6,*]

[1]Committee on Clinical Pharmacology and Pharmacogenomics, The University of Chicago, Chicago, IL 60637, USA

[2]Section of Hematology/Oncology, The University of Chicago, Chicago, IL 60637, USA

[3]Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL 60637, USA

[4]Department of Surgery, The University of Chicago, Chicago, IL 60637, USA

[5]Department of Health Studies, The University of Chicago, Chicago, IL 60637, USA

[6]Center for Personalized Therapeutics, The University of Chicago, Chicago, IL 60637, USA

[7]RIKEN Center for Genomic Medicine, Yokohama, Japan

Abstract

Cytochrome P450 2D6 (cytochrome P450, family 2, subfamily D, polypeptide 6, or CYP2D6), a highly polymorphic drug metabolizing enzyme, is involved in the metabolism of one quarter of the most commonly prescribed medications. Here, we have applied multiple genotyping methods and Sanger sequencing to assign precise and reproducible *CYP2D6* genotypes, including copy numbers, for 48 HapMap samples. Furthermore, by analyzing a set of 50 human liver microsomes using endoxifen formation from N-desmethyl-tamoxifen as the phenotype of interest, we observed a significant positive correlation between CYP2D6 genotype-assigned activity score and endoxifen formation rate ($r_s = 0.68$ by Rank correlation test, $P = 5.3 \times 10^{-8}$), which corroborated the genotype-phenotype prediction derived from our genotyping methodologies. In the future, these 48 publicly available HapMap samples characterized by multiple substantiated *CYP2D6* genotyping platforms could serve as a reference resource for assay development, validation, quality control, and proficiency testing for other *CYP2D6* genotyping projects, and for programs pursuing clinical pharmacogenomic testing implementation.

*To whom correspondence should be addressed at: 5841 S. Maryland Avenue, MC 2115, Chicago, IL 60637, USA. Tel.: +1 773 702 7564, Fax: +1 773 702 3163; podonnel@medicine.bsd.uchicago.edu.

Keywords

CYP2D6; genotyping; pharmacogenomics clinical implementation; sequencing

Introduction

Cytochrome P450 2D6 (cytochrome P450, family 2, subfamily D, polypeptide 6, or CYP2D6) is involved in the phase I metabolism of approximately one quarter of the most commonly prescribed medications, including β-blockers, antiarrhythmics, opioids, anticancer drugs, and a number of antidepressant and antipsychotic agents[1, 2]. For example, tamoxifen, which is widely used for the treatment and prevention of recurrence of hormone receptor-positive breast cancer, is also one of the most investigated CYP2D6 substrates since CYP2D6 plays a significant role in the formation of its active metabolites – 4-hydroxytamoxifen and endoxifen.

The *CYP2D6* gene, which is located on chromosome 22q13.1, is highly polymorphic, and to date over 100 defined allele variants have been reported (http://www.cypalleles.ki.se/cyp2d6.htm). Some genetic variants in *CYP2D6* can significantly affect its enzymatic activity, and four CYP2D6 phenotypes are commonly defined: poor metabolizer (PM), intermediate metabolizer (IM), extensive metabolizer (EM), and ultrarapid metabolizer (UM)[3]. Variant alleles of *CYP2D6* consist of singlenucleotide polymorphisms (SNPs), small insertions and deletions, gene rearrangements, hybrid genes, and copy number variations (CNV) including deletion or duplications/multiplications of the entire gene[4]. The deletion of the entire *CYP2D6* gene (*5*) leads to the absence of enzyme activity (i.e., PM phenotype), whereas duplications or multiplications of the functional gene result in overexpression of CYP2D6 (i.e., UM phenotype). In addition, the presence of two highly homologous pseudogenes, *CYP2D7* and *CYP2D8*, in physical proximity to *CYP2D6* has made accurate *CYP2D6* genotyping even more difficult[5]. Furthermore, there are important ethnic differences in the frequency of functional *CYP2D6* alleles. For example, 5–10% of Caucasian populations have a PM phenotype by carrying two null alleles (especially *3*, *4*, *5*, or *6* among others), while another 1–2% of Caucasians are UMs who typically carry a duplicated/multiplied *CYP2D6*2xN* gene [6, 7]. In contrast, the majority of Asians are categorized as IMs due to the high frequency of a reduced function allele*CYP2D6*10* (e.g., ~40% in east Asian population) while PM or UM phenotypes are fairly uncommon[8].

Due to the clinical significance of the medications metabolized by CYP2D6, it is critical, especially in a clinical setting, to obtain an accurate estimation of CYP2D6 metabolic activity based on determination of *CYP2D6* genotype. Several genotyping/sequencing platforms have been developed to discern *CYP2D6* genotypes in an effort to improve the accuracy of phenotype prediction for patients; however, there are very few well characterized and validated *CYP2D6* reference materials available for public access. Previously, Pratt et al.[9] reported highly valuable information on a set of DNA reference materials but there were technical (platform) limitations to the scope of their work.

Here, we have applied multiple genotyping methodologies and Sanger sequencing method to assign precise and reproducible *CYP2D6* genotypes, including gene copy number, for 48

HapMap samples from European and Yoruba ancestry. Wedid not include Asian populations in our current study due to the presence of already established, large and well-characterized reference samples for Japanese and Han Chinese populations[10, 11]

One of the genotyping methods we applied in this study is the invader assay coupled with multiplex PCR, also known as multiplexPCR-basedreal-timeinvaderassay (mPCR-RETINA), whichhas been described as a highly accurate, high-throughputSNP genotyping method[12]. RETINA monitors the fluorescence intensity of eachvariation locus in real time, and is able to detect variant asymmetries caused by CNV in heterozygous individuals[13].Furthermore, we also confirmed the utility of our genotyping methods for CYP2D6 metabolic activity prediction by analyzing a set of 50 human liver microsomes using endoxifen formation from N-desmethyl-tamoxifen as the phenotype of interest.

Materials and Methods

Genomic DNA Samples and Human Liver Samples

To comprise the proposed reference set, 48 genomic DNA samples (26of European unrelated ancestry and 22 of Yorubaunrelated ancestry) from the International HapMap project were purchased from Coriell Cell Repositories (Camden, NJ).

Forty-four (44) Caucasian and six (6) African American human livers were donated by healthy human subjects via the Liver Tissue Procurement and Distribution System and the Cooperative Human Tissue Network with approval of the respective Institutional Review Boards.The use of these livers was deemed exempt from ethical review by the Institutional Review Board at The University of Chicago.Human liver microsomes were prepared as previously described[14]. Protein concentrations were measured using the Qubit® protein assay kit (Thermo Fisher Scientific, Pittsburgh, PA). DNA was isolated from 20 mg of liver tissue using the Blood and Cell culture mini kit (Qiagen, Valencia, CA)following the manufacturer's method for tissue samples.

Multiple PCR-based Real-Time Invader Assay (mPCR-RETINA)

For the RETINA assay, the entire *CYP2D6* gene region was first amplified by a single triplex PCR reaction by using Invader triplex PCR primer pairs 1, 2, and 3 to generate three shorter, non-overlapping fragments. Primer pair 1 amplifies exons 1 and 2, primer pair 2 amplifies exons 3 to 6, and primer pair 3 amplifies exons 7 to 9. Genomic DNA concentration was detected by Nanodrop spectrophotometer (Thermo Scientific, Logan, UT). Ten (10) ng of genomic DNA was used for each sample and the Takara Ex Taq HS PCR system (Clontech Laboratories, Inc.) was applied according to the manufacturer's instructions; the PCR conditions were as follows: initiation at 95°C for 2 min, 35 cycles of 98°C for 10 s and 68°C for 3 min, and termination at 72°C for 2 min. For*CYP2D6*41*(2988 G>A) detection, a separate short amplicon PCR reaction was performed because the triplex RETINA products do not cover that position. Invader PCR primer pairs for *41* were used and the PCR conditions were as follows: initiation at 95°C for 2 min, 35 cycles of 98°C for 10 s and 68°C for 30 s, and termination at 72°C for 1 min.PCR amplified DNA samples

were then diluted (1:10 dilution) and used as templates for the chosen Invader assays (see Supplementary Table 1).

The 29 variants of *CYP2D6* detected by RETINA were as follows (Table 1): 31 G>A, 77 G>A, 100 C>T,124 G>A, 137_138 insT, 883 G>C, 1023 C>T,1659 G>A, 1707 delT, 1716 G>A, 1758 G>T, 1863_1864ins(TTTCGCCCC)$_2$,1846 G>A, 1973_1974 insG, 2291 G>A, 2539_2542 del AACT, 2549 delA, 2573_2574 insC, 2587_2590 delGACT, 2615_2617 delAAG, 2850 C>T, 2935 A>C, 2950 G>C, 2988 G>A, 3183 G>A, 3201 C>T, 3259_3260 insGT,4125_4133 dup GTGCCCACT, gene conversion to *CYP2D7* in exon 9.

Fluorescence resonance energy transfer (FRET) probes labeled with FAM or Yakima Yellow were purchased from Third Wave Technologies (now Hologic, Inc., Bedford, MA). Rox dye (6-carboxy-X-rhodamine) used for the normalization of reporter signals was purchased from Sigma-Aldrich (St. Louis, MO). In each reaction, 0.75 μl of 10x signal buffer (Third Wave Technologies, now Hologic, Inc., Bedford, MA), 0.5 μl of FRET/ROX (10:3) mixture, 0.25 μl cleavase 2.0 (Third Wave Technologies, now Hologic, Inc., Bedford, MA), 0.375 μl 20x allele and invader probe mixture, 5.65 μl of water, and 2.5 μl diluted PCR product (1:10 diluted) were mixed and incubated at 98°C for 5 min and 65°C for 5 min. Genotyping results were processed and analyzed by ViiA 7 Real-time PCR system (Life Technologies, Carlsbad, CA). All primer and probe sequences are listed in Supplementary Table 1.

TaqMan® Drug Metabolism Genotyping Assays for CYP2D6

Nine (9) *CYP2D6* TaqMan® drug metabolism genotyping assays (assay IDs:C_34816116_20, C_27102425_10, C_32407229_60, C_32407240_80, C_27102431_D0, C_2222771_40, C_11484460_40, C_27102444_80, and C_34816113_20) were tested according to the manufacturer's protocol. Per reaction, 10 ng of genomic DNA was used along with 2x TaqMan Universal PCR Master Mix (Life Technologies, Carlsbad, CA). The PCR conditions were initiation at 95°C for 10 min and 50 cycles of 92°C for 15 s and 60°C for 90 s.

The following alleles were included in the genotyping assays: 2988 G>A (*CYP2D6*41*), 2850 C>T (*CYP2D6*2*), 2615_2617 delAAG (*CYP2D6*9*), 1863_1864 ins(TTTCGCCCC)$_2$ (*CYP2D6*40*), 1846 G>A (*CYP2D6*4*), 1023 C>T (*CYP2D6*17*), 100 C>T (*CYP2D6*10*), 31 G>A (*CYP2D6*35*), and 3189 G>A (*CYP2D6*29*).

Direct Capillary Sequencing/Sanger Sequencing

The *CYP2D6* gene for each sample was first amplified using two specific primers (DPKup and DPKlow; Supplementary Table 1) to generate a 5kb *CYP2D6* region (Chromosome: 22; 42522040 – 42527140) by the Takara LA Taq PCR system (Clontech Laboratories, Inc., Mountain View, CA). Ten (10) ng of genomic DNA was used in the volume of 20 μl reaction and the PCR conditions were as follows: initiation at 95°C for 2 min, 30 cycles of 98°C for 10 s and 68°C for 30 s, and termination at 72°C for 7 min.The amplified *CYP2D6* samples were then purified by Agencourt AMPure XP Beads (Beckman Coulter, Beverly, MA) and subjected to direct DNA sequencing by using 14 *CYP2D6*-specific sequencing

primers (Supplementary Table 1).The sequencing PCR protocol of BigDye Terminator version 3.1 (Life Technologies, Carlsbad, CA)was applied and samples were sequenced using the3500xl Genetic Analyzer (Life Technologies, Carlsbad, CA).

Fourteen sequencing fragments of each sample were aligned by using DNA Baser sequence assembly software and compared with the *CYP2D6* references (GenBank Accession Number M33388 and AY545216). *CYP2D6* nonmenclature standards (www.cypalleles.ki.se/cyp2d6.htm) were applied to define the haplotype or "*" star variant alleles.

CYP2D6*5 detection by Long-Range PCR

Two different sets of PCR primers (D1/D2 and 13/24) were adopted for detection of *CYP2D6*5* by long-range PCR[15].Long-range PCR products were analyzed by 1.0% agarose gel electrophoresis. The presence of two fragments, 6.0 kb (by primers D1 and D2) and 3.5 kb (by primers 13 and 24)in length, respectively,was indicative of the presence of the deletion (*CYP2D6*5* allele). Ten (10) ng of genomic DNA was used for each sample and the Takara LA Taq PCR system was applied according to the manufacturer's instructions. The PCR conditions were as follows: initiation at 95°C for 2 min, 35 cycles of 94°C for 30 s, 66°C for 30 s, and 68°C for 5 min, and termination at 72°C for 7 min. All primer sequences are listed in Supplementary Table 1.

Copy Number Assays by TaqMan® Real-time PCR

To access *CYP2D6* gene copy number, three TaqMan® real-time PCR assays targeting different regions of the *CYP2D6* gene were used. All TaqMan® assays and reagents were purchased from Life Technologies (Carlsbad, CA), including three commercial quantitative TaqMan® copy number assays (assay IDs: Hs00010001_cn targeting exon 9, Hs04502391_cn targeting intron 6, and Hs04083572_cn targeting intron 2) and one TaqMan® copy number reference assay, RNase P, human (assay ID: 4403326). All assays were performed in triplicate along with an internal control RNaseP assay according to the manufacturer's protocol directly using genomic DNAs. Briefly, 10 ng of genomic DNA was used in the volume of 10 μl reaction and the PCR conditions were as follows: hold at 95°C for 10 min, 40 cycles of 95°C for 15 s and 60°C for 60 sby ViiA 7 Real-time PCR system (Life Technologies, Carlsbad, CA). Relative quantification of *CYP2D6* copy number was performed using CopyCaller Software (Life Technologies, Carlsbad, CA) following the comparative delta-delta threshold cycle (C_T) method. Each assay was repeated twice.

Comparison with Data from the 1000 Genomes Project

The 1000 Genomes data for *CYP2D6* were extracted in PLINK binary format[16] from the 1000 Genomes Project sequence data frozen from 23 Nov 2010 (low-coverage whole-genome) and 21 May 2011 (high-coverage exome).

CYP2D6 Phenotyping Assay in Human Liver Microsomes

Endoxifen formation was investigated in 50 human liver microsomes by using *N*-desmethyl-tamoxifen as substrate. *N*-desmethyl-tamoxifen and endoxifen were obtained from Toronto Research Chemicals (Toronto, Canada). Verapamil and HPLC-grade methanol were

purchased from Fisher Scientific Company LLC (Hanover Park, IL). Triethylammonium phosphate (1 M solution) was purchased from Sigma-Aldrich (St. Louis, MO). NADPH regenerating system solutions A and B were obtained from BD Biosciences (Bedford, MA). Experiments were performed under low-light conditions to avoid photodegradation of compounds. Pilot experiments were performed with pooled human liver microsomes to optimize incubation conditions with respect to time (range: 10–120 min), microsomal protein (32–625 μg) and endoxifen concentration (1–50 μM). Incubations with individual human liver microsomes were performed under linear conditions and contained 62.5 μg of human liver microsomal protein, 20 μM substrate and 100 mM potassium phosphate buffer (pH 7.4) in a final volume of 250 μl. Reactions were initiated by addition of an NADPH regenerating system (1.3 mM NADP$^+$, 3.3 mM glucose-6-phosphate, 3.3 mM MgCl$_2$ and 0.4 U/ml glucose-6-phosphate dehydrogenase), conducted at 37°C for 25 min, and terminated with 100 μl of cold acetonitrile. After addition of 4.8 μM verapamil (internal standard), samples were centrifuged (20,817 rcf for 15 min at 4°C); aliquots (100 μl) were analyzed by high performance liquid chromatography (HPLC) equipped with a Peltier sample cooler set at 4°C to prevent sample degradation. Endoxifen formation was measured as previously published [17] with the following modification in the mobile phase (5 mM triethylammonium phosphate (TEAP) and acetonitrile): 65/35 (v/v) from 0–20 min and 50/50 (v/v) from 20.1–60 min. Retention times were 8, 15 and 30 min for verapamil, endoxifen and *N*-desmethyl-tamoxifen, respectively. For enhancing sensitivity through photochemical derivatization, a post-column photochemical reactor enhancement detection system (PHRED) was added. Fluorescence detection was performed with an excitation λ= 256 nm and emission λ= 380 nm and formation rates were expressed as pmol/min/mg protein.The quantitation limit was 0.4 pmol/min/mg protein with intra-day precision (%CV) and accuracy=4% and 113%, and inter-day precision (%CV) and accuracy=13% and 102%.

Statistical Analysis

Square root transformation was chosen to achieve approximate normality of endoxifen rate using Box-Cox approach [18, 19]. The cases with zero endoxifen formation rate were shifted to the smallest non-zero value divided by ten to allow log transformation. The optimal transformation was not sensitive to different values of this shift. Rank correlation test was performed to show robustness of the results to modeling assumptions. Calculations were performed using the statistical software R [20].

Results

Genotyping by RETINA and TaqMan® Drug Metabolism Genotyping Assays

We first used two different genotyping assays for the genotyping of the 48 HapMap samples. We tested pre-designed RETINA assays for 29 common variations of *CYP2D6* with 29 potential detectable *CYP2D6* alleles(Table 1, Supplementary Table 1). Twelve (12) loci were polymorphic among the 48 samples: 31 samples (65%)had a 2850C>T SNP, 13 (27%) had a100C>T SNP, 10 (21%) had a 1846G>A SNP, 9 (19%) had a 1023 C>T SNP, 9 (19%) had a 2988 G>A SNP, 3 (6%) had a 31 G>A SNP, 2 (4%) had a 1716 G>A SNP, 2 (4%) had a 1659 G>A SNP, 2 (4%) had a 3183 G>A SNP, 1 (2%) had a 77 G>A SNP, 1 (2%) had a 1863_1864ins(TTTCGCCCC)$_2$ variant, and 1 (2%) had a 2615_2617delAAG

variant. Representative data of the 1846G>A and 100C>T SNP assays of the RETINA system for all 48 samples are shown in Figure 1A. Two samples, #36 and #45, had clear asymmetry patterns via RETINA, suggesting the presence of CNV.

To verify these polymorphic loci, we applied a second, independent genotyping method – TaqMan® Drug Metabolism CYP2D6 Genotyping assays. Among the 12 RETINA-detected polymorphic loci in our samples, however, only 9 were availablecommerciallyvia TaqMan® (31 G >A, 100C>T, 1023 C>T, 1846G>A, 1863_1864ins(TTTCGCCCC)$_2$, 2615_2617 delAAG,2850C>T, 2988G>A, and 3183 G>A).We tested those 9TaqMan® Drug Metabolism Genotyping assays and found no discrepancies in variant calls between the two platforms. Representative data of the 1846G>A and 100 C>T SNPs from the TaqMan® Drug Metabolism assays are shown in Figure 1B. Similar to RETINA, we also observed clear asymmetry patterns for samples #36 and #45. Surprisingly, sample #2 (*CYP2D6*2/*4*) also showed asymmetry pattern in TaqMan® Drug Metabolism Genotyping assay100C>T, but not in the RETINA 100C>T assay.Further CNV methods were applied to carefully detect exact copy number in those samples.

CNV detection by TaqMan® Real-time PCR and Long-range PCR

We first used the TaqMan® real-time PCR for copy number calculation, and three different assays were applied due to the complex nature of *CYP2D6*. These three assays targeted different regions of the *CYP2D6* gene:intron 2 (Int2), intron 6 (Int6), and exon 9 (Ex9). Among the 48 samples, 6 (13%) had one copy of *CYP2D6*, 37 (77%) had two copies, 4 (8%) had three copies, and 1 (2%) had four copies; all three assays showed concordant results (Figure 2). Interestingly, Hapmap samples #36 and #45 were found to have three and four copies of *CYP2D6*, respectively, which was consistent with previous asymmetric genotyping results (Figure 1). However, sample #2 did not carry multiple copies of *CYP2D6*; thus, we concluded that TaqMan® Metabolism assays were not as accurate as RETINA in determining copy number variation.

To confirm the *CYP2D6*5* allele (whole gene deletion), we performed long-range PCR, which has been widely used as a standard method since it was first published by Steen *et al.*[21]. Here, we applied two different sets of primers (D1/D2 and 13/24) to avoid miscalls for *CYP2D6*5*[15]. All six samples showing one copy by TaqMan® copy number assays were able to show *CYP2D6*5*-specific bands by long-range PCR (3.5 kb and 6 kb; Supplementary Figure 1), but none of the other samples showed the same patterns. These data verified the ability to consistently detect the presence of the *CYP2D6*5* allele in our reference set by either of the two independent methods.

Direct Capillary Sequencing/Sanger Sequencing

To validate the accuracy of genotyping results by the prior two methods, we next performed Sanger sequencing on the 48 Hapmap samples. Overall, there were fully concordant results between RETINA assays and Sanger sequencing, confirming the accuracy of RETINA. However, TaqMan® Drug Metabolism CYP2D6 Genotyping assaysincludeda more limited number of detectable alleles. Thus, genotypes of Hapmap samples #18, #39, #47 could not

be called accurately according to the TaqMan® Drug Metabolism CYP2D6 Genotyping assaysthat we used.

Combining exact copy number from TaqMan® assays, the *5 allele status by long-range PCR, and the variant discrimination plots from RETINA and Sanger sequencing, we were able to provide composite allelic calls for all of the samples (Table 2). The duplication or multiplication assignment of samples with multiple copies of *CYP2D6* was based on the asymmetric RETINA clustering.

Comparison with the data from the 1000 Genomes Project

To further evaluate our genotyping results, we next examined our data of the HapMap samples to *CYP2D6* sequencing results from The 1000 Genomes Project, which was generated from a next-generation sequencing platform[22]. Among 29variations covered by the RETINA system in our study, only 11(31 G>A, 100 C>T, 883 G>C, 1023 C>T, 1707 delT, 1758 G>T, 1846 G>A,2549 delA, , 2615_2617 del AAG, , 2850 C>T, and 2988 G>A) were found within the 1000 Genomes data in 47 of 48 samples for comparison (data for sample #4, NA07346, was not available). Copy number information about *CYP2D6* was not available for any of the samples from 1000 Genomes data.

Two variant discrepancies were found. In sample #28, 2549 delA SNP was identified by the 1000 Genomes data, but was not detected by genotyping or Sanger sequencing by us. Similarly, in sample #41, 31 G>A heterozygous SNP was identified by the 1000 Genomes data, however, our assays showed a wild-type genotype at that locus (Table 2).

Genotype and Phenotype Correlation using Human Liver Samples

To further investigate the applicability of our genotyping methods for phenotype prediction, we analyzed a set of 50 human samples – liver microsomes from healthy Caucasian and African American donors. Due to the prior convincing data of the genotyping assays, we only applied the RETINA assay system to genotype these samples.

Among the 50 samples (Table 3), 12 different *CYP2D6* loci were identified to be polymorphic: 29 samples (58%)had a 2850C>T SNP, 16 samples (32%) had a 1846G>A SNP, 15 samples (30%) had a100C>T SNP, 14 samples (28%) had a 2988 G>A SNP, 5 samples (10%) had a 31 G>A SNP, 4 samples (8%) had a 1707delT variant, 3 samples (6%) had a 2291G>A SNP, 2 samples (4%) had a 1023 C>T SNP, 2 samples (4%) had a variant of gene conversion with *CYP2D7* in exon 9, 1 sample (2%) had a 1659 G>A SNP, 1 sample (2%) had a 3183 G>A SNP, and 1 sample (2%) had a 2549 delA variant. Four (4) out of 12 loci detected in these liver samples were not found in the Hapmap samples: 2549 delA (*3), 1707 del T (*6), exon 9 conversion with *CYP2D7* (*36 or *4N), and 2291 G>A (*59).

For copy number detection, both TaqMan® real-time PCR and long-range PCR were used(Supplementary Figure 2); 40 samples were found to have two copies of *CYP2D6*, 5 samples (10%) were found to have three copies. Six (6) samples (12%) were found to carry the *CYP2D6*5* allele, one of which was copy-neutral (liver sample #32L), so only 5 samples were found to have one copy of *CYP2D6*.

Interestingly, liver sample #32L had a *CYP2D6*5* allele indicated by long-range PCR but also had two copies of *CYP2D6* by TaqMan® copy number assay, indicating the possibility of two genotypes: (a) *CYP2D6*1×2/*5* with duplication of *CYP2D6* on one chromosome and *CYP2D6* gene deletion on the other; or (b) *CYP2D6*1-*5/*1* with one *CYP2D6* copy and gene deletion on the same chromosome, and one *CYP2D6* copy on the other chromosome (the latter case is hypothetical as it hasnot to our knowledge yet been described).

Separately, two Caucasian liver samples #8L and #35L, both of which had three copies of *CYP2D6* gene, were found to carry a heterozygous 1846 G>A SNP, a heterozygous 100 C>T SNP, and a heterozygous variant of the *CYP2D6* gene conversion with *CYP2D7* in exon 9. Gene conversion with *CYP2D7* in exon 9 is the 'key' allele for*CYP2D6*36*, however, it can also be found in a sub-variant*CYP2D6*4*(*4N*) that is exon 9 conversion-positive[23]. Thus, allelic pattern of these samples could represent one of several actual genotypes:*1/*4Nx2, *1×2/*4N, *4/*36×2,* or *4×2/*36*. Due to the rare frequency of *36* gene arrangements in Caucasians, both samples were hypothezied to carry the *CYP2D6*4N* subvariant with a genotype of either *CYP2D6*1/*4Nx2* or *CYP2D6*1×2/*4N*. To distinguish the exact genotype, we then performed Sanger sequencing on these two samples. Since the height of A allele (*4*) was twice the height of G allele (*1*) at position 1846 (Supplementary Figure 3), and since there was aprevious report that *4N* was only found in duplication arrangement[23], we concluded that these two samples carried two copies of *4* and one copy of *1*, resulting in a variant call of*CYP2D6*1/*4Nx2*.

Next, based on the genotype information, we calculated the CYP2D6 activity scores[24] and assigned predicted metabolic status of all the samples (Table 3). We then examined correlation between our assigned CYP2D6 activity score (AS) and the phenotype data, which is the rate of endoxifen formation from *N*-desmethyl-tamoxifen metabolism mainly through CYP2D6 (Figure 3). For assigned AS groups of 0, 0.5, 1.0, 1.5, and 2.0, the rate of endoxifen formation (mean ± SD) was 0.20 ± 0.18 (n= 7), 0.33 ± 0.22 (n=4), 2.01 ± 0.83 (n=14), 1.72± 1.08 (n=14), and 3.74 ± 1.47 (n=10) pmol/min/mg *N*-desmethyl-tamoxifen, respectively (for the AS group of 2.5, there is only one sample with a value of 3.32 pmol/min/mg). Thus, with an increase of predicted AS, an elevated endoxifen formation rate was observed; this positive correlation (r_s = 0.72 by square root transformation, P = 4.2 ×10^{-9}; r_s = 0.68 by Rank correlation test, P = 5.3 × 10^{-8}) indicated high concordance of genotype-phenotype prediction based on our genotyping methodologies.

However, we observethat several samples (Figure 3, in red circles) had unexpectedly lowactual metabolic activity compared to their assigned AS. For example, sample #28L with a *CYP2D6*2/*4* genotype and assigned AS of 1.0, sample #19L with a *CYP2D6*2/*59* genotype and assigned AS of 1.5, #30L with a *CYP2D6*2/*59* genotype and assigned AS of 1.5, #38L with a *CYP2D6*35/*41* genotype and assigned AS of 1.5, and #15L with a *CYP2D6*1/*35* genotypeand assigned AS of 2.0 all showed somewhat lower endoxifen formation ratesthan predicted by the genotype or AS.Sanger sequencing was performed on these samples, however, nomissing variations were found.

Expectedly, we also did not observe overall differences in endoxifen formation rates between theAS 1.5 group andthe AS 1.0 group because these two AS groups are considered clinically indistinguishable. Thus, they both belong to theextensive metabolizer (EM) group[24].

Additionally, for the two Caucasian liver samples #8L and #35L, the endoxifen formation rates were 1.25 ± 0.11 and 3.01 ± 0.14 pmol/min/mg *N*-desmethyl-tamoxifen, respectively, which were in the range of expected endoxifen formation rate from AS group of 1.0 (based on genotype **1/*4Nx2*) rather than AS group of 0 (based on genotype **4×2/*36* or **4/ *36×2*). These concordant phenotypes, again, corroborated the correct assigned genotypes for these two samples.

Discussion

Accurate *CYP2D6* genotyping including assessment of copy number has been historically challenging because of the structural complexity of the gene. Comparisons of different genotyping platforms often leads to lack of consensus agreement about genotype calls [9, 25, 26], making verification difficult and the critical assignment of phenotype for this important drug metabolizing enzyme problematic. In this paper, we have described the application of multiple *CYP2D6* genotyping/sequencing approaches by analyzing 48 publicly-available genomic DNA samples and 50 human liver samples in order to develop a well-characterized reference set of samples with consistent *CYP2D6* genotypes verified by multiple methods. Overall, concordant results were observed between the multiple methods, including via the typical gold standard, Sanger sequencing. We conclude that our results, and the development of the characterized samples as a possible reference set, could permit: a) the application of two relatively easy-to-perform genotyping methods combined with copy number assays for accurate use in characterizing this complex gene in other future projects; and b) the availability of a known group of samples to serve as *CYP2D6* reference samples for other laboratories to use when developing or validating individual *CYP2D6* genotyping assays.

In a previous study[9] that applied five different commercially available platforms (Roche Amplichip, AutoGenomics INFINITI, Luminex, ParagonDx, and LDT SNaPShot) to characterize 107 genomic DNA samples, genotype discrepancies were often found between different platforms for *CYP2D6*, largely related to the variability in allelic coverage and allele definition. For example, AutoGenomics INFINITI and Luminex xTag, which covered only 15 and 13 *CYP2D6* alleles, respectively, are not designed to identify *CYP2D6*35* because they donot detect 31 G>A. Roche Amplichip had the best allele coverage among the platforms used in this prior study, however,it doesnot include the defining *CYP2D6*41* SNP 2988G>A and thus may misclassify some *CYP2D6*2* alleles as *CYP2D6*41*. Additionally, in that study, discrepantgenotypes (which did exist between different methods) were not adjudicated/confirmed by Sanger or other sequencing methods. Our study, in contrast, has severalsignificant advantages. First, one of our genotyping methods - RETINA - covers29 of the mostfrequent alleles of the *CYP2D6* gene, all of which have known correlations with enzymatic activity for clinical application. By combining copy number assays (including*CYP2D6*5* detection) with RETINA, our approach also delivers significantly

better allelic coverage than those *CYP2D6* genetic tests cleared by the U.S. Food and Drug Administration (FDA)–the Roche AmpliChip and Luminex xTAG CYP2D6 kit. For example, when compared with the most comprehensive FDA-cleared commercial panel – Roche AmpliChip – our platform can detect ten additional loci (124 G>A [*CYP2D6*12*], 4142_4133 dupGTGCCCACT [*CYP2D6 *18*], 2573_2574 insC [*CYP2D6*21*], 2587_2590 delGACT [*CYP2D6*38*],77 G>A[*CYP2D6*43*],1716 G>A [*CYP2D6*45*], 77 G>A and 1716 G>A [*CYP2D6*46*],3259_3260 insGT [*CYP2D6*42*], 3201 C>T [*CYP2D6*56*], and 2291 G>A [*CYP2D6*59*]). For the RETINA assays for which positive signals were not observed in our samples, mosthave been successfully detected with RETINA in the previous report of a large Asian sample set[10].Secondly, we used the most updated nomenclature(http://www.cypalleles.ki.se/cyp2d6.htm) to generate allele calls (Table 1). Third, Sanger sequencing was applied to confirm the accuracy of our genotyping data. Further, it should also be noted that some of the commercially available genetic testing platforms do not include an assessment of *CYP2D6* copy number (e.g. xTAG® CYP2D6 from Luminex, DMET™ from Affymetrix) and thus cannot adequately assign accurate diplotypes for *CYP2D6*. In contrast, our method was not only comprehensive by including 29 clinically actionable alleles, but also accurate by delivering results of exact copy number including *5 detection. It is acknowledged that *CYP2D6*13*-like *CYP2D7/2D6* hybrid genes[27], which arefairly rare (~0.1–0.2%) in Caucasian and African populations[28], were not included in our RETINA panel. All ten different *CYP2D6*13* subvariants share a *CYP2D7*-derived exon 1 with the detrimental T-insertion, but differ in respect to the region in which *CYP2D7* switches to *CYP2D6*.There is no diagnostic SNP for *CYP2D6*13*, and thus we argue that it would not be optimal to try to assess this variant via an Invader or Taqman technology, but rather assessment would best be done by sequencing the entire gene region.

We also validated our genotyping approaches using a panel of human liver microsomes by comparing the assigned genotypes with CYP2D6 activity scores in this well-validated *in vitro* system.This genotype to phenotype prediction is extremely important for future pharmacogenomic clinical implementation. For those samples that had unexpectedly low metabolizing activitycompared to assigned AS based on *CYP2D6* genotype, this may be due to some degree of heterogeneity within individual collected human liver samples, or potentially to epigenetic modifications resulting in disassociation of genotype with phenotype in these cases. We were unable to find evidence of sample degradation as a potential cause.

In our study, the TaqMan® copy number assay was straightforward and found to be highly accurate by showing concordant results from three assays targeting different regions of the *CYP2D6* gene. When combined with long-range PCR, it can without ambiguity allow assignment of copy number including detection of the *5 allele. We used two different primer sets to avoid miscalling for *CYP2D6*5* since a novel structure of *CYP2D6* has been previously reported that can generate false positive results by using only one primer set [15]. However, from the standpoint of generating fully precise genomic data, the TaqMan® copy number assay alone is not sufficient to replace long-range PCR since it cannot detect the *5 allele in copy-neutral cases, like liver sample #32L. However, the clinically relevant predicted enzymatic activity score of such deletion-copy-neutral samples would not be

affected by this shortcoming. In this example, if copy-neutral-deletion/duplication were unappreciated via the TaqMan® copy number assay alone, the assigned genotype would be (*1/*1), which has the same activity score and phenotype (2.0; EM) as the more precise genotype result from long-range PCR (*1/*1/*5; 2.0; EM). This suggests the feasibility of replacing the tedious long-range PCR method that requires agarose gel electrophoresisby the relatively easy TaqMan® copy number assay in a clinical setting if prediction of the enzymatic activity score or phenotype is the desired endpoint.

Although we observed concordant data among different analytical methods, we did find two allele discrepancies between our results and The 1000 Genomes data. We suspect they are most likely to be representative of sequencing errors of the next-generation sequencing (NGS) platform(s) applied in The 1000 Genomes Project, perhaps due to limited coverage depth for the *CYP2D6* region. It is also possible that NGS sequences may not be correctly aligned. Furthermore, considering the lack of copy number information, the 1000 Genomes data may not serve as a suitable reference for highly polymorphic genes like *CYP2D6*.

In summary, we have validated the application of two genotyping methods in combination with the TaqMan® copy number assay for accurate use in characterizing *CYP2D6* in other future projects. Additionally, we have developed a reference set of 48 publicly available HapMap samples now accurately characterized on a genomic level via our consistent genotyping methodologies. These samples will hopefully enable the development, validation, quality control, and proficiency testing for other *CYP2D6* genotyping projects, including those potentially attempting implementation of *CYP2D6* genotyping in a Clinical Laboratory Improvement Amendments (CLIA) setting. In fact, the above findings have indeed permitted the assessment and planned delivery of *CYP2D6* genotype and phenotype results for a large cohort of patients currently participating in an institutional pharmacogenomics clinical implementation project - The 1200 Patients Project [29].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Zhou SF. Polymorphism of human cytochrome P450 2D6 and its clinical significance: part II. Clinical pharmacokinetics. 2009; 48(12):761–804. [PubMed: 19902987]

2. Zhou SF. Polymorphism of human cytochrome P450 2D6 and its clinical significance: Part I. Clinical pharmacokinetics. 2009; 48(11):689–723. [PubMed: 19817501]

3. Zanger UM, Raimundo S, Eichelbaum M. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. Naunyn-Schmiedeberg's archives of pharmacology. 2004; 369(1):23–37.

4. Gaedigk A. Complexities of CYP2D6 gene analysis and interpretation. International review of psychiatry. 2013; 25(5):534–553. [PubMed: 24151800]

5. Kramer WE, Walker DL, O'Kane DJ, Mrazek DA, Fisher PK, Dukek BA, et al. CYP2D6: novel genomic structures and alleles. Pharmacogenet Genomics. 2009; 19(10):813–822. [PubMed: 19741566]

6. Broly F, Gaedigk A, Heim M, Eichelbaum M, Morike K, Meyer UA. Debrisoquine/sparteine hydroxylation genotype and phenotype: analysis of common mutations and alleles of CYP2D6 in a European population. DNA Cell Biol. 1991; 10(8):545–558. [PubMed: 1681816]

7. Sachse C, Brockmoller J, Hildebrand M, Muller K, Roots I. Correctness of prediction of the CYP2D6 phenotype confirmed by genotyping 47 intermediate and poor metabolizers of debrisoquine. Pharmacogenetics. 1998; 8(2):181–185. [PubMed: 10022755]

8. Bradford LD. CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. Pharmacogenomics. 2002; 3(2):229–243. [PubMed: 11972444]

9. Pratt VM, Zehnbauer B, Wilson JA, Baak R, Babic N, Bettinotti M, et al. Characterization of 107 genomic DNA reference materials for CYP2D6, CYP2C19, CYP2C9, VKORC1, and UGT1A1: a GeT-RM and Association for Molecular Pathology collaborative project. J Mol Diagn. 2010; 12(6): 835–846. [PubMed: 20889555]

10. Hosono N, Kato M, Kiyotani K, Mushiroda T, Takata S, Sato H, et al. CYP2D6 genotyping for functional-gene dosage analysis by allele copy number detection. Clin Chem. 2009; 55(8):1546–1554. [PubMed: 19541866]

11. Qian JC, Xu XM, Hu GX, Dai DP, Xu RA, Hu LM, et al. Genetic variations of human CYP2D6 in the Chinese Han population. Pharmacogenomics. 2013; 14(14):1731–1743. [PubMed: 24192122]

12. A haplotype map of the human genome. Nature. 2005; 437(7063):1299–1320. [PubMed: 16255080]

13. Hosono N, Kubo M, Tsuchiya Y, Sato H, Kitamoto T, Saito S, et al. Multiplex PCR-based real-time invader assay (mPCR-RETINA): a novel SNP-based method for detecting allelic asymmetries within copy number variation regions. Hum Mutat. 2008; 29(1):182–189. [PubMed: 17879341]

14. Ramirez J, Liu W, Mirkov S, Desai AA, Chen P, Das S, et al. Lack of association between common polymorphisms in UGT1A9 and gene expression and activity. Drug Metab Dispos. 2007; 35(12):2149–2153. [PubMed: 17761781]

15. Fukuda T, Maune H, Ikenaga Y, Naohara M, Fukuda K, Azuma J. Novel structure of the CYP2D6 gene that confuses genotyping for the CYP2D6*5 allele. Drug Metab Pharmacokinet. 2005; 20(5): 345–350. [PubMed: 16272752]

16. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81(3):559–575. [PubMed: 17701901]

17. Antunes MV, Rosa DD, Viana Tdos S, Andreolla H, Fontanive TO, Linden R. Sensitive HPLC-PDA determination of tamoxifen and its metabolites N-desmethyltamoxifen, 4-hydroxytamoxifen and endoxifen in human plasma. J Pharm Biomed Anal. 2013; 76:13–20. [PubMed: 23291438]

18. Box GEPaC, D. R. An analysis of transformations (with discussion). Journal of the Royal Statistical Society. 1964; (B26):211–252.

19. Venables WNaR, B. D.. Modern Applied Statistics with S. Fourth edition. Springer; 2002.

20. Tea RDC. R: a language and environment for statistical computing. 2011 R Foundation for Statistical Computing.

21. Steen VM, Andreassen OA, Daly AK, Tefre T, Borresen AL, Idle JR, et al. Detection of the poor metabolizer-associated CYP2D6(D) gene deletion allele by long-PCR technology. Pharmacogenetics. 1995; 5(4):215–223. [PubMed: 8528268]

22. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491(7422):56–65. [PubMed: 23128226]

23. Gaedigk A, Bradford LD, Alander SW, Leeder JS. CYP2D6*36 gene arrangements within the cyp2d6 locus: association of CYP2D6*36 with poor metabolizer status. Drug Metab Dispos. 2006; 34(4):563–569. [PubMed: 16415111]

24. Crews KR, Gaedigk A, Dunnenberger HM, Klein TE, Shen DD, Callaghan JT, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype. Clin Pharmacol Ther. 2012; 91(2):321–326. [PubMed: 22205192]

25. Heller T, Kirchheiner J, Armstrong VW, Luthe H, Tzvetkov M, Brockmoller J, et al. AmpliChip CYP450 GeneChip: a new gene chip that allows rapid and accurate CYP2D6 genotyping. Therapeutic drug monitoring. 2006; 28(5):673–677. [PubMed: 17038884]

26. Kim J, Lee SY, Lee KA. Copy number variation and gene rearrangements in CYP2D6 genotyping using multiplex ligation-dependent probe amplification in Koreans. Pharmacogenomics. 2012; 13(8):963–973. [PubMed: 22676200]

27. Sim SC, Daly AK, Gaedigk A. CYP2D6 update: revised nomenclature for CYP2D7/2D6 hybrid genes. Pharmacogenet Genomics. 2012; 22(9):692–694. [PubMed: 22868903]

28. Gaedigk A, Fuhr U, Johnson C, Berard LA, Bradford D, Leeder JS. CYP2D7-2D6 hybrid tandems: identification of novel CYP2D6 duplication arrangements and implications for phenotype prediction. Pharmacogenomics. 2010; 11(1):43–53. [PubMed: 20017671]

29. O'Donnell PH, Bush A, Spitz J, Danahey K, Saner D, Das S, et al. The 1200 patients project: creating a new medical model system for clinical implementation of pharmacogenomics. Clin Pharmacol Ther. 2012; 92(4):446–449. [PubMed: 22929923]
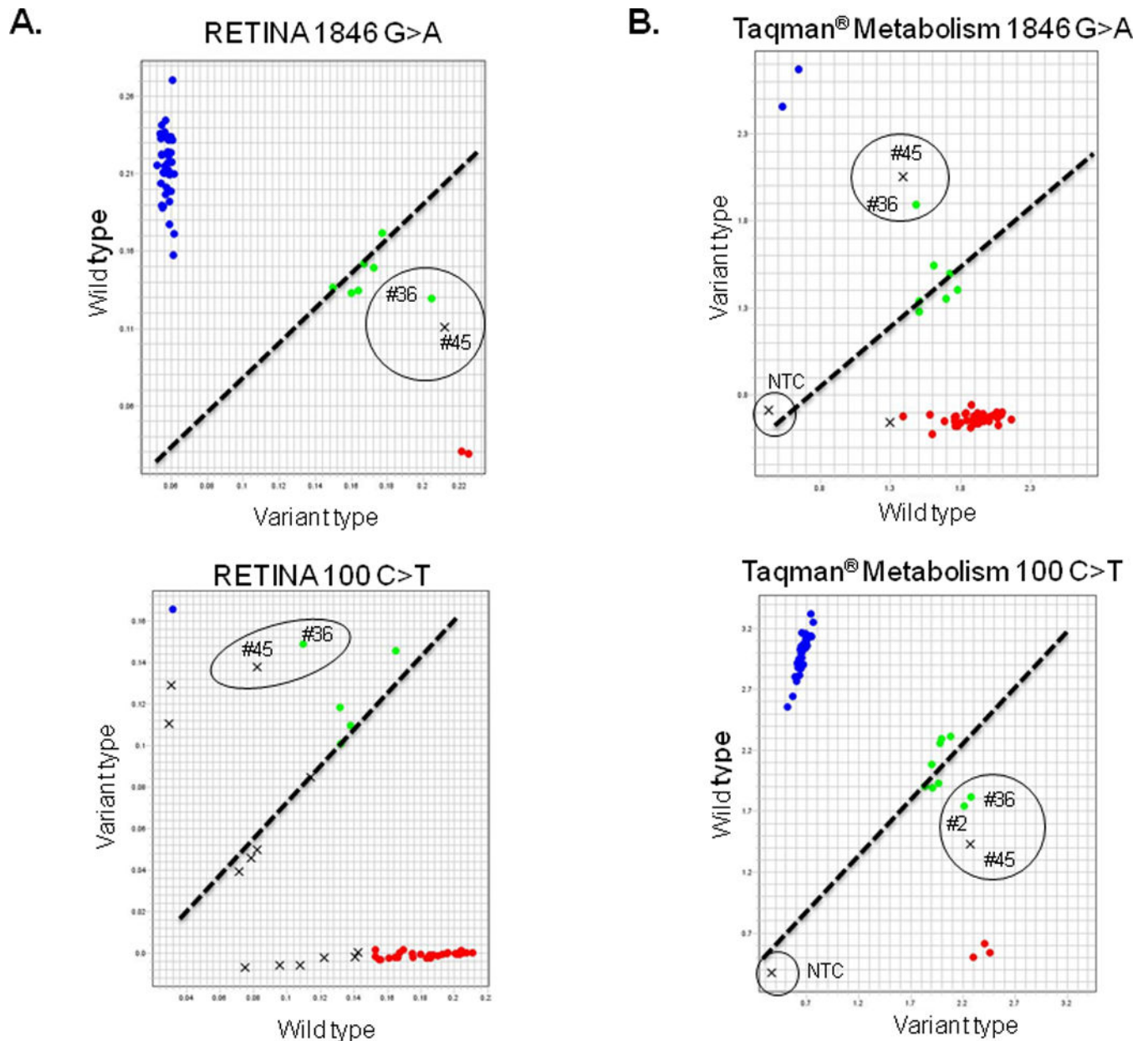
Figure 1.
Determination of *CYP2D6* SNPs by two genotyping methods. Variation discrimination plots of 1846 G>A (upper panels; *4*) and 100 C>T (lower panels: *10/*4*) by mPCR RETINA (A) and TaqMan® Drug Metabolism Assay (B). NTC = no template control. The presentation of a dot or an X for a sample is dependent on the automatic assignment from the real-time PCR genotyping software based on the signal intensity of that particularsample.
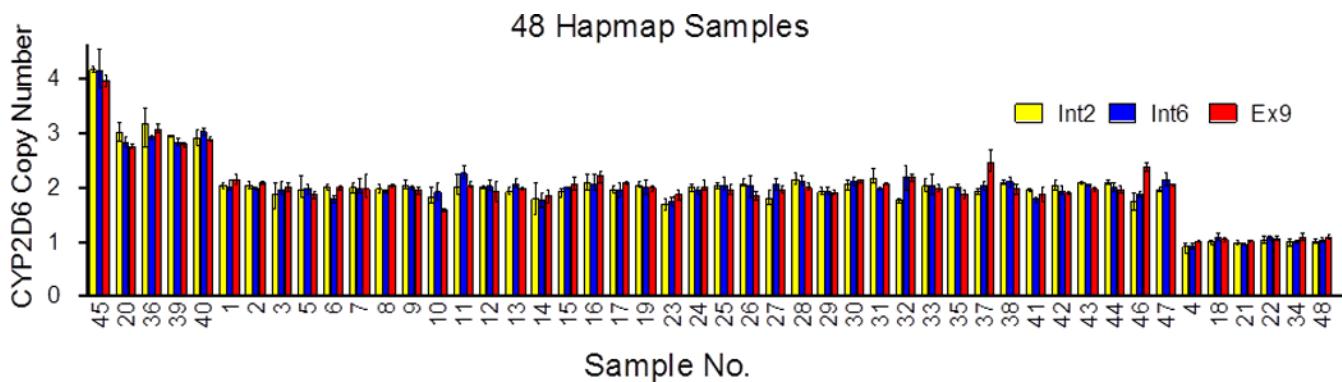
Figure 2.
Estimation of *CYP2D6* gene copy number in 48 HapMap samples by three TaqMan[®] copy number assays. Comparison of *CYP2D6* gene copy number assignments.*CYP2D6* copy numbers (y-axis) were estimated by three assays that targeted intron 2 (Int2; yellow), intron 6 (In6; blue), and exon 9 (Ex9; red). Each sample was assayed in triplicate for each assay and the values are the means of the detected *CYP2D6* copy numbers with the bars representing the maximum and minimum estimates.
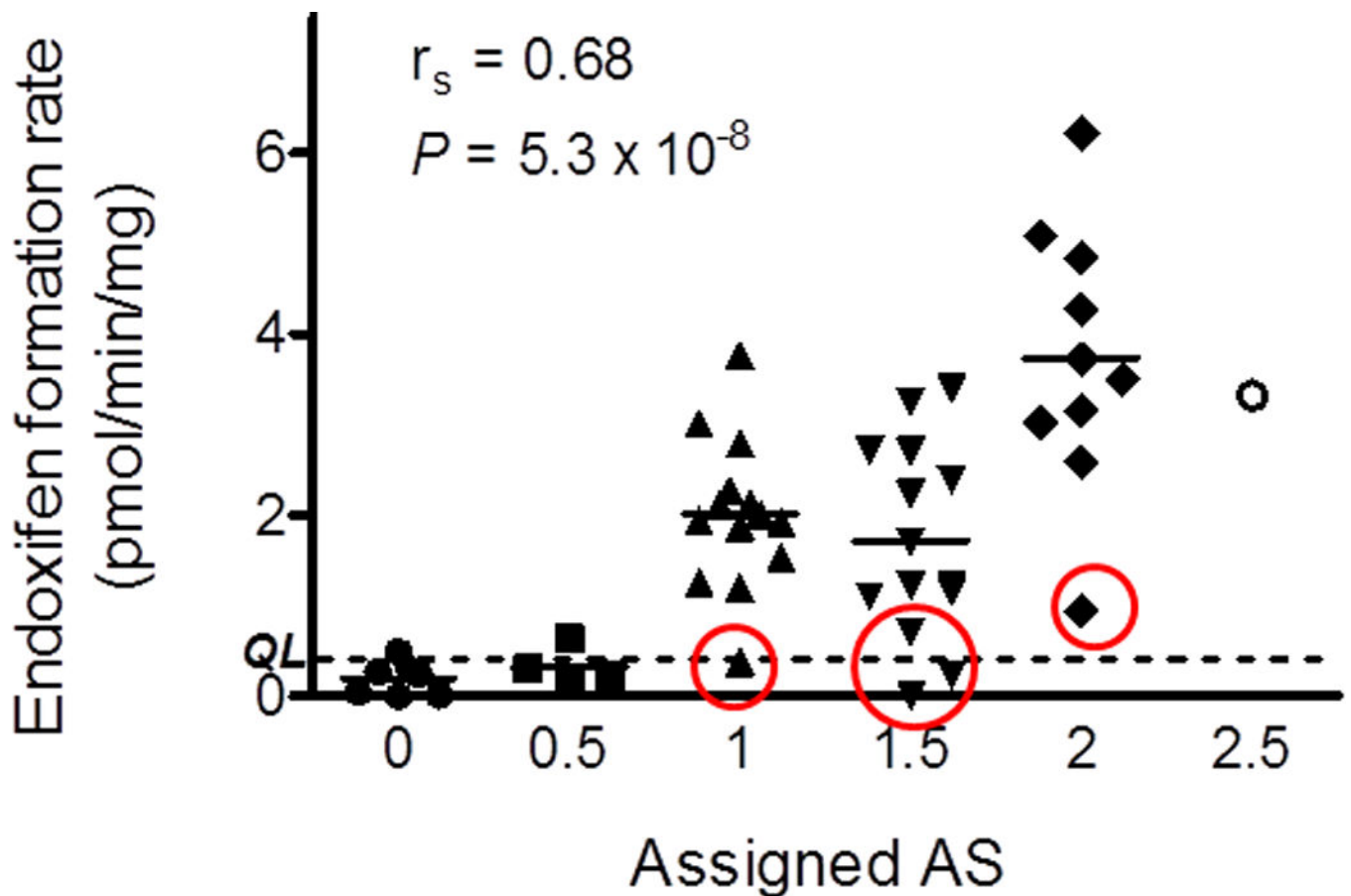
Figure 3.
Correlation between genotype-assigned activity score of CYP2D6 and endoxifen formation
rate from N-desmethyl-tamoxifen in 50 liver samples. For 50 human liver microsomes
samples, activity scores (AS) of CYP2D6 were assigned based on genotyping calls, and
endoxifen formation was investigated by using *N*-desmethyl-tamoxifen as substrate as
described in Materials and Methods. A strong positive correlation between increased
predicted AS and elevated endoxifen formation rate was observed ($r_s = 0.72$ by square root
transformation, $P = 4.2 \times 10^{-9}$; $r_s = 0.68$ by Rank correlation test, $P = 5.3 \times 10^{-8}$).
The quantitation limit (QL) was 0.4 pmol/min/mg protein for endoxifen formation rate.

Table 1

Sequence Variants Utilized to Detect *CYP2D6* alleles.

CYP2D6 Alleles	Variant[†]
*2	**2850 C>T**
*3	**2549 del A**
*4	100 C>T, **1846 G>A, (exon 9 converstion with *CYP2D7* for *CYP2D6* *4N)**
*6	**1707 del T**
*7	**2935 A>C**
*8	**1758 G>T**, 2850 C>T
*9	**2615_2617delAAG**
*10	**100 C>T**
*11	**883 G>C**, 2850 C>T
*12	**124 G>A**, 2850 C>T
*15	**137_138insT**
*17	**1023 C>T**, 2850 C>T
*18	**4125_4133dupGTGCCCACT**
*19	**2539_2542 delAACT**, 2850 C>T
*20	**1973_1974 insG**, 2850 C>T
*21	**2573_2574 insC**, 2850 C>T
*29	**1659 G>A, 3183 G>A**, 2850 C>T
*35	**31 G>A**, 2850 C>T
*36	100 C>T, **exon 9 conversion with *CYP2D7***
*38	**2587_2590 del GACT**
*40	1023 C>T, **1863_1864 ins(TTTCGCCCC)$_2$**, 2850 C>T
*41	2850 C>T, **2988 G>A**
*42	2850 C>T, **3259_3260 insGT**
*43	**77 G>A**
*44	**2950 G>C**
*45	**1716 G>A**, 2850 C>T
*46	**77 G>A, 1716 G>A**, 2850 C>T
*56	2850 C>T, **3201 C>T**
*59	**2291 G>A**, 2850 C>T

[†]Defining variant in bold; additional sequence variations may be present (for details, please refer to *CYP2D6* nomenclature website).

Table 2

CYP2D6 genotyping results of HamMap samples from multiple methodologies.

Sample #	Hapmap #	Ethnicity	RETINA	Taqman® Drug Metabolism Genotyping Assays	Sanger sequencing	LR PCR	CNV (TaqMan)	Final Integrated Genotype
1	NA06994	Caucasian	*1/*1	*1/*1	*1/*1	-	2	*1/*1
2	NA07037	Caucasian	*2/*4	*2/*4	*2/*4	-	2	*2/*4
3	NA07048	Caucasian	*1/*4	*1/*4	*1/*4	-	2	*1/*4
4	NA07346	Caucasian	*2/*2	*2/*2	*2/*2	*5	1	*2/*5
5	NA11933	Caucasian	*35/*41	*35/*41	*35/*41	-	2	*35/*41
6	NA11993	Caucasian	*1/*9	*1/*9	*1/*9	-	2	*1/*9
7	NA12045	Caucasian	*1/*41	*1/*41	*1/41	-	2	*1/*41
8	NA12058	Caucasian	*2/*41	*2/*41	*2/*41	-	2	*2/*41
9	NA12287	Caucasian	*41/*41	*41/*41	*41/*41	-	2	*41/*41
10	NA12399	Caucasian	*1/*1	*1/*1	*1/*1	-	2	*1/*1
11	NA12718	Caucasian	*1/*1	*1/*1	*1/*1	-	2	*1/*1
12	NA12750	Caucasian	*2/*2	*2/*2	*2/*2	-	2	*2/*2
13	NA12751	Caucasian	*1/*2	*1/*2	*1/*2	-	2	*1/*2
14	NA12775	Caucasian	*1/*10	*1/*10	*1/*10	-	2	*1/*10
15	NA12814	Caucasian	*2/*41	*2/*41	*2/*41	-	2	*2/*41
16	NA12827	Caucasian	*2/*35	*2/*35	*2/*35	-	2	*2/*35
17	NA18501	Yoruba	*1/*17	*1/*17	*1/*17	-	2	*1/*17
18	NA18502	Yoruba	*45/*45	*2/*2 (no *45 assay available)	*45/*45	*5	1	*5/*45
19	NA19129	Yoruba	*17/*17	*17/*17	*17/*17	-	2	*17/*17
20	NA19137	Yoruba	*2/*17	*2/*17	*2/*17	-	3	*2×2/*17
21	NA19200	Yoruba	*1/*1	*1/*1	*1/*1	*5	1	*1/*5
22	NA19209	Yoruba	*17/*17	*17/*17	*17/*17	*5	1	*5/*17
23	NA06984	Caucasian	*4/*4	*4/*4	*4/*4	-	2	*4/*4
24	NA10851	Caucasian	*1/*4	*1/*4	*1/*4	-	2	*1/*4
25	NA11830	Caucasian	*1/*4	*1/*4	*1/*4	-	2	*1/*4
26	NA11843	Caucasian	*1/*41	*1/*41	*1/*41	-	2	*1/*41

Sample #	Hapmap #	Ethnicity	RETINA	Taqman® Drug Metabolism Genotyping Assays	Sanger sequencing	LR PCR	CNV (TaqMan)	Final Integrated Genotype
27	NA11893	Caucasian	*1/*2	*1/*2	*1/*2	-	2	*1/*2
28	NA11920	Caucasian	*1/*4	*1/*4	*1/*4	-	2	*1/*4
29	NA12282	Caucasian	*4/*4	*4/*4	*4/*4	-	2	*4/*4
30	NA12347	Caucasian	*1/*41	*1/*41	*1/*41	-	2	*1/*41
31	NA12843	Caucasian	*1/*35	*1/*35	*1/*35	-	2	*1/*35
32	NA12889	Caucasian	*4/*41	*4/*41	*4/*41	-	2	*4/*41
33	NA18867	Yoruba	*2/*10	*2/*10	*2/*10	-	2	*2/*10
34	NA18910	Yoruba	*2/*2	*2/*2	*2/*2	*5	1	*2/*5
35	NA18917	Yoruba	*1/*17	*1/*17	*1/*17	-	2	*1/*17
36	NA18924	Yoruba	*2/*4	*2/*4	*2/*4	-	3	*2/*4×2
37	NA19114	Yoruba	*1/*1	*1/*1	*1/*1	-	2	*1/*1
38	NA19117	Yoruba	*1/*40	*1/*40	*1/*40	-	2	*1/*40
39	NA19152	Yoruba	*29/*43	*1/*29 (no *43 assay available)	*29/*43	-	3	*29/*43×2
40	NA19171	Yoruba	*2/*41	*2/*41	*2/*41	-	3	*2×2/*41
41	NA19222	Yoruba	*1/*1	*1/*1	*1/*1	-	2	*1/*1
42	NA19225	Yoruba	*17/*17	*17/*17	*17/*17	-	2	*17/*17
43	NA19235	Yoruba	*1/*17	*1/*17	*1/*17	-	2	*1/*17
44	NA19257	Yoruba	*1/*1	*1/*1	*1/*1	-	2	*1/*1
45	NA19175	Yoruba	*1/*4	*1/*4	*1/*4	-	4	*1/*4×3
46	NA19147	Yoruba	*17/*29	*17/*29	*17/*29	-	2	*17/*29
47	NA18505	Yoruba	*1/*45	*1/*2 (no *45 assay available)	*1/*45	-	2	*1/*45
48	NA18517	Yoruba	*10/*10	*10/*10	*10/*10	*5	1	*5/*10

RETINA = Polymerase chain reaction-based real-time invader assay ; LR-PCR = long range PCR; CNV = copy number variation

Table 3

CYP2D6 genotyping results and predicted enzymatic activities of 50 human liver samples.

Sample Number	Source ID	Ethnicity	RETINA	LR PCR	CNV (TaqMan®)	Final Integrated Genotype	Predicted Activity Score	Predicted Metabolic Status
1L	N/A	AA	*4/*4	-	3	*4×2/*4	0	PM
2L	UC9208	AA	*1/*29	-	2	*1/*29	1.5	EM
3L	N/A	Caucasian	*4/*41	-	2	*4/*41	0.5	IM
4L	N/A	Caucasian	*1/*1	-	2	*1/*1	2.0	EM
5L	N/A	Caucasian	*2/*41	-	2	*2/*41	1.5	EM
6L	UC9305	Caucasian	*1/*4	-	2	*1/*4	1.0	EM
7L	N/A	Caucasian	*1/*6	-	2	*1/*6	1.0	EM
8L	N/A	Caucasian	*1/*4	-	3	*1/*4×2	1.0	EM
9L	N/A	AA	*4/*17	-	3	*4×2/*17	0.5	IM
10L	UC9306	Caucasian	*1/*1	-	2	*1/*1	2.0	EM
11L	UC9307	Caucasian	*1/*2	-	2	*1/*2	2.0	EM
12L	UC9308	AA	*2/*4	-	2	*2/*4	1.0	EM
13L	UC9310	Caucasian	*2/*41	-	3	*2×2/*41	2.5	UM
14L	N/A	Caucasian	*4/*35	-	2	*4/*35	1.0	EM
15L	UC9406	Caucasian	*1/*35	-	2	*1/*35	2.0	EM
16L	UC9504	Caucasian	*6/*6	*5	1	*5/*6	0	PM
17L	UC9506	Caucasian	*1/*41	-	2	*1/*41	1.5	EM
18L	UC9507	Caucasian	*1/*3	-	2	*1/*3	1.0	EM
19L	HH761	Caucasian	*2/*59	-	2	*2/*59	1.5	EM
20L	HH768	AA	*1/*17	-	2	*1/*17	1.5	EM
21L	HH659	Caucasian	*1/*41	-	2	*1/*41	1.5	EM
22L	HH745	Caucasian	*2/*59	-	2	*2/*59	1.5	EM
23L	HH775	Caucasian	*4/*4	-	2	*4/*4	0	PM
24L	HH776	Caucasian	*1/*1	-	2	*1/*1	2.0	EM
25L	HH785	Caucasian	*1/*2	-	2	*1/*2	2.0	EM
26L	HH789	Caucasian	*2/*41	-	2	*2/*41	1.5	EM

Sample Number	Source ID	Ethnicity	RETINA	LR PCR	CNV (TaqMan®)	Final Integrated Genotype	Predicted Activity Score	Predicted Metabolic Status
27L	HH790	Caucasian	*1/*2	-	2	*1/*2	2.0	EM
28L	HH792	Caucasian	*2/*4	-	2	*2/*4	1.0	EM
29L	HH806	Caucasian	*2/*41	-	2	*2/*41	1.5	EM
30L	HH824	Caucasian	*2/*59	-	2	*2/*59	1.5	EM
31L	HH830	Caucasian	*41/*41	-	2	*41/*41	1.0	EM
32L	HH839	Caucasian	*1/*1	*5	2	*1×2/*5	2.0	EM
33L	HH840	Caucasian	*4/*4	-	2	*4/*4	0	PM
34L	HH841	Caucasian	*2/*2	*5	1	*2/*5	1.0	EM
35L	HH844	Caucasian	*4/*1	-	3	*1/*4×2	1.0	EM
36L	HH848	Caucasian	*1/*1	-	2	*1/*1	2.0	EM
37L	HH850	Caucasian	*1/*1	-	2	*1/*1	2.0	EM
38L	HH861	Caucasian	*35/*41	-	2	*35/*41	1.5	EM
39L	HH864	Caucasian	*41/*41	*5	1	*5/*41	0.5	IM
40L	HH870	Caucasian	*2/*41	-	2	*2/*41	1.5	EM
41L	HH873	Caucasian	*4/*6	-	2	*4/*6	0	PM
42L	HH874	AA	*1/*41	-	2	*1/*41	1.5	EM
43L	N/A	Caucasian	*4/*4	-	2	*4/*4	0	PM
44L	N/A	Caucasian	*4/*4	-	2	*4/*4	0	PM
45L	N/A	Caucasian	*35/*41	-	2	*35/*41	1.5	EM
46L	N/A	Caucasian	*1/*1	*5	1	*1/*5	1.0	EM
47L	N/A	Caucasian	*1/*1	*5	1	*1/*5	1.0	EM
48L	N/A	Caucasian	*1/*4	-	2	*1/*4	1.0	EM
49L	N/A	Caucasian	*4/*41	-	2	*4/*41	0.5	IM
50L	N/A	Caucasian	*6/*35	-	2	*6/*35	1.0	EM

RETINA = Polymerase chain reaction-based real-time invader assay ; LR-PCR = long-range PCR; CNV = copy number variation; PM = poor metabolizer; IM = intermediate metabolizer; EM = extensive metabolizer; UM = ultrarapid metabolizer; AA = African American; N/A = not available.