Establishment of CYP2D6 reference samples by multiple validated genotyping platforms

H Fang1,2, X Liu2, J Ramírez2, N Choudhury2, M Kubo3, HK Im3, A Konkashbaev5, NJ Cox1,5,6, MJ Ratain1,2,6, Y Nakamura1,2,6,7 and PH O'Donnell1,2,6

INTRODUCTION

Cytochrome P450 2D6 (cytochrome P450, family 2, subfamily D, polypeptide 6 (CYP2D6)) is involved in the phase I metabolism of approximately one-quarter of the most commonly prescribed medications, including β-blockers, antiarrhythmics, opioids, anticonvulsants, 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 as CYP2D6 has 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 > 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, extensive metabolizer (EM), and ultrarapid metabolizer (UM).3 Variant alleles of CYP2D6 consist of single-nucleotide polymorphisms (SNPs), small insertions and deletions, gene rearrangements, hybrid genes and copy number variations (CNVs), 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 (that is, PM phenotype), whereas duplications or multiplications of the functional gene result in overexpression of CYP2D6 (that is, 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 UM who typically carry a duplicated/multiplied CYP2D6*2 × N gene.5,6 In contrast, the majority of Asians are categorized as intermediate metabolizers owing to the high frequency of a reduced function allele CYP2D6*10 (for example, ~40% in east Asian population) while PM or UM phenotypes are fairly uncommon.6 Because of 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.7 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. We did not include Asian populations in our current study owing to the presence of already established, large and well-characterized reference samples for Japanese and Han Chinese populations.8,9

1 Committee on Clinical Pharmacology and Pharmacogenomics, The University of Chicago, Chicago, IL, USA; 2Section of Hematology/Oncology, The University of Chicago, Chicago, IL, USA; 3RIKEN Center for Genomic Medicine, Yokohama, Japan; 4Department of Health Studies, The University of Chicago, Chicago, IL, USA; 5Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, USA; 6Center for Personalized Therapeutics, The University of Chicago, Chicago, IL, USA and 7Department of Surgery, The University of Chicago, Chicago, IL, USA. Correspondence: Dr P O’Donnell, Department of Medicine, The University of Chicago, 5841 S. Maryland Avenue, Mail Code 2115, Chicago, IL 60637, USA.
E-mail: podonnel@medicine.bsd.uchicago.edu
Received 18 December 2013; revised 30 April 2014; accepted 22 May 2014; published online 1 July 2014
One of the genotyping methods we applied in this study is the invader assay coupled with multiplex PCR, also known as multiplex PCR-based real-time invader assay (mPCR-RETINA), which has been described as a highly accurate, high-throughput SNP genotyping method. RETINA monitors the fluorescence intensity of each variation locus in real time and is able to detect variant asymmetries caused by CNV in heterozygous individuals. 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 (26 of the European unrelated ancestry and 22 of the Yoruba unrelated ancestry) from the International HapMap project were purchased from Coriell Cell Repositories (Camden, NJ, USA).

CYP2D6*41 was included (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, Chicago, IL, USA. Human liver microsomes were prepared as previously described. Protein concentrations were measured using the Qubit protein assay kit (Thermo Fisher Scientific, Pittsburgh, PA, USA). DNA was isolated from 20 mg of liver tissue using the Blood and Cell culture mini kit (Qiagen, Valencia, CA, USA) following the manufacturer’s method for tissue samples.

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 3–6 and primer pair 3 amplifies exons 7–9. Genomic DNA concentration was detected by Nanodrop spectrophotometer (Thermo Scientific, Logan, UT, USA). Ten nanograms of genomic DNA was used for each sample and the Takara Ex Taq HS PCR system (Clontech Laboratories, Inc., Mountain View, CA, USA) 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 amplification 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 S1).

The 29 variants of CYP2D6 detected by RETINA are as follows (Table 1): G>A, 77 C>T, 124 G>A, 137_138 insT, 883 G>C, 1023 C>T, 1659 G>A, 1707 delG, 1716 G>A, 1758 G>T, 1863_1864 ins(TTGCGCC), 1846 G>A, 1973_1974 insG, 2291 G>A, 2539_2542 delAACT, 2549 delA, 2573_2574 insC, 2587_2590 delGACT, 2615_2617 delAAG, 2850 C>T, 2935 A>C, 2950 G>C, 2988 G>A, A313 G>A, 3201 C>T, 3259_3260 insGT, 4125_4133 dup GTGCCCACT, and gene conversion.

Fluorescence resonance energy transfer (FRET) probes labeled with TAMRA or Yakima Yellow were purchased from Third Wave Technologies (now Hologic, Inc., Bedford, MA, USA). Rox dye (6-carboxy-X-rhodamine) used for the normalization of reporter signals was purchased from Sigma-Aldrich (St Louis, MO, USA). In each reaction, 0.75 μl of 10 x signal buffer (Third Wave Technologies, now Hologic, Inc.), 0.5 μl of FRET/Rox (10:3) mixture, 0.25 μl cleavage 2.0 (Third Wave Technologies, now Hologic, Inc.), 0.375 μl 20 x 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, USA). All primer and probe sequences are listed in Supplementary Table S1.

TaQMan drug metabolism genotyping assays for CYP2D6
Nine CYP2D6 TaqMan drug metabolism genotyping assays (assays IDs: C_34816116_20, C_27102425_10, C_32407229_60, C_32407240_80, C_27102431_D0, C_2222771_40, C_2222846_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 2 x TaqMan Universal PCR Master Mix (Life Technologies). The PCR conditions were as follows: 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(TTGCGCC), (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 S1) to generate a 5-kb CYP2D6 region (Chromosome: 22; 42522040–42527140) by the Takara LA Taq PCR system (Clontech Laboratories, Inc.). Ten nanograms 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, USA) and subjected to direct DNA sequencing by using 14 CYP2D6-specific sequencing primers (Supplementary Table S1).

The sequencing PCR protocol of BigDye Terminator version 3.1 (Life Technologies) was applied, and the samples were sequenced using the 3500xl Genetic Analyzer (Life Technologies). Fourteen sequencing fragments of each sample were aligned by using DNA Baser sequence assembly software and compared with the CYP2D6 references (GenBank Accession numbers M333388 and AY545216). CYP2D6 nonnculare standards (http://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. 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 nanograms 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 S1.

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, including three commercial quantitative TaqMan copy number assays (assay IDs: Hs00010001_cn targeting exon 9, Hs04502931_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 s by Viia 7 Real-time PCR system (Life Technologies). Relative quantification of CYP2D6 copy number was performed using the CopyCaller Software (Life Technologies) following the comparative delta-delta threshold cycle (ΔΔCt) 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 from the 1000 Genomes Project sequence data frozen between 23 November 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-desethyl-tamoxifen as substrate. N-desethyl-tamoxifen and endoxifen were obtained from Toronto Research Chemicals (Toronto, Canada). Verapamil and high-performance liquid chromatography (HPLC)-grade methanol were purchased from Fisher Scientific Company LLC (Hanover Park, IL, USA). Triethylammonium phosphate (1M solution) was purchased from Sigma-Aldrich (St Louis, MO, USA). NADPH regenerating system solutions A and B were obtained from BD Biosciences (Bedford, MA, USA). 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: 10–20 min and 50/50 (v/v) from 20.1 to 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 was added. Fluorescence detection was performed with an excitation wavelength λ = 356 nm and emission wavelength λ = 380 nm, and formation rates were expressed as pmol·min⁻¹·mg⁻¹ protein. The quantitation limit was 0.4 pmol·min⁻¹·mg⁻¹ protein with intra-day precision (%CV) and accuracy = 4% and 113%, respectively, and inter-day precision (%CV) and accuracy = 13% and 102%, respectively.

Statistical analysis

Square root transformation was chosen to achieve approximate normality of endoxifen rate using the Box–Cox approach.18,19 The cases with zero endoxifen formation rate were shifted to the smallest non-zero value divided by 10 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 predesigned RETINA assays for 29 common variations of CYP2D6 with 29 potential detectable CYP2D6 alleles (Table 1, Supplementary Table S1). Twelve loci were polymorphic among the 48 samples: 31 samples (65%) had a 2850 C>T SNP, 13 (27%) had a 100 C>T SNP, 10 (21%) had a 1846 G>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_1864 ins(TTTCGCCCC)_2 SNP, and 1 (2%) had a 2615_2617 delAAAG variant. Representative data of the 1846 G>A and 100 C>T SNP assays of the RETINA system for all the 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 available commercially via TaqMan (31 G>A, 100 C>T, 1023 C>T, 1846 G>A, 1863_1864 ins(TTTCGCCCC)_2, 2615_2617 delAAAG, 2850 C>T, 2988 G>A, 3183 G>A). We tested those 9 TaqMan Drug Metabolism Genotyping assays and found no discrepancies in variant calls between the two platforms. Representative data of the 1846 G>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 assay 100 C>T but not in the RETINA 100 C>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 (In2t), intron 6 (In6t), 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 CNV.

To confirm the CYP2D6*5 allele (whole gene deletion), we performed long-range PCR, which has been widely used as a standard method as it was first published by Steen et al.23 Here we
applied two different sets of primers (D1/D2 and 13/24) to avoid miscalls for CYP2D6*5. 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 and 6 kb; Supplementary Figure S1), 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 previous 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 assays included a more limited number of detectable alleles. Thus, genotypes of Hapmap samples #18, #39 and #47 could not be called accurately according to the TaqMan Drug Metabolism Assay (b) that 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. Among 29 variations covered by the RETINA system in our study, only 11 (31G>4, 100 C>T, 883G>4, 1707 delT, 1758G>T, 1846G>A, 2549 delA, 2615_2617 del AAG, 2850 C>T and 2988 G>A) were found within the 1000 Genomes data in 47 of the 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 the 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. Because of the previous 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 a 100C>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 out of the 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 S2); 40 samples were found to have two copies of CYP2D6, and 5 samples (10%) were found to have three copies. Six samples (12%) were found to carry the CYP2D6*5 allele, one of which was copy-neutral (liver sample #32L), so only five 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*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 one chromosome, and one CYP2D6 copy on the other chromosome (the latter case is hypothetical as it has not, 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, all the patterns of these samples could represent one of the several actual genotypes: *1/4N*2, *1*2/4N, *4/36*2, or *4*2/36. Because of the rare frequency of *36 gene arrangements in Caucasians, both samples were hypothesized to carry the CYP2D6*4N subvariant with a genotype of either CYP2D6*1/*4N*2 or CYP2D6*1/*4N. To distinguish the exact genotype, we then performed Sanger sequencing on these two samples. As the height of A allele (*4) was twice the height of G allele (*1) at position 1846 (Supplementary Figure S3), and as there was a previous 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/*4N*2.

Next, based on the genotype information, we calculated the CYP2D6 activity scores (ASs) and assigned predicted metabolic status of all the samples (Table 3). We then examined correlation between our assigned CYP2D6 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 ± s.d.) 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⁻¹ mg⁻¹ N-desmethyl-tamoxifen, respectively (for the AS group of 2.5, there is only one sample with a value of 3.32 pmol min⁻¹ mg⁻¹). Thus, with an increase of predicted AS, an elevated endoxifen formation rate was observed; this positive correlation (r = 0.72 by square root transformation, P = 4.2 × 10⁻¹⁰; r = 0.68 by Rank correlation test, P = 5.3 × 10⁻⁹) indicated high concordance of genotype-phenotype prediction based on our genotyping methodologies.

However, we did observe that several samples (Figure 3, in red circles) had unexpectedly low actual metabolic activity compared with 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, #15L with a CYP2D6*1/*35 genotype and assigned AS of 2.0 all showed somewhat lower endoxifen formation rates than predicted by the genotype or AS. Sanger sequencing was performed on these samples; however, no missing variations were found.

Expectedly, we also did not observe overall differences in endoxifen formation rates between the AS 1.5 group and the AS 1.0 group, because these two AS groups are considered clinically indistinguishable. Thus they both belong to the 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⁻¹ mg⁻¹ N-desmethyl-tamoxifen, respectively, which were in the range of expected endoxifen formation rate from the AS group of 1.0 (based on genotype *1/4N*2) rather than the AS group of 0 (based on genotype *4/2/36). 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 lead 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 study9 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

### Table 2. CYP2D6 genotyping results of HamMap samples from multiple methodologies

| Sample no. | Hapmap no. | 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             | —      | —           | *1/*1                    |
| 2         | NA07037    | Caucasian | *2/*4  | *2/*4                                    | *2/*4             | —      | —           | *2/*4                    |
| 3         | NA07048    | Caucasian | *1/*4  | *1/*4                                    | *1/*4             | —      | —           | *1/*4                    |
| 4         | NA07346    | Caucasian | *2/*2  | *2/*2                                    | *2/*2             | *5     | 1           | *2/*5                    |
| 5         | NA11933    | Caucasian | *35/*41| *35/*41                                   | *35/*41           | —      | —           | *35/*41                  |
| 6         | NA11993    | Caucasian | *1/*9  | *1/*9                                    | *1/*9             | —      | —           | *1/*9                    |
| 7         | NA12045    | Caucasian | *1/*41 | *1/*41                                   | *1/*41           | —      | —           | *1/*41                   |
| 8         | NA12058    | Caucasian | *2/*41 | *2/*41                                   | *2/*41           | —      | —           | *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) | — | 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           | *2x2/*17                 |
| 21        | NA19200    | Yoruba    | *1/*1  | *1/*1                                    | *1/*1             | 5      | 1           | *1/*5                    |
| 22        | NA19209    | Yoruba    | *17/*17| *17/*17                                  | *17/*17           | 5      | —           | *17/*5                   |
| 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                   |
| 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           | *2x4/*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                                   | *29/*43          | 3      | 3           | *29/43x2                 |
| 40        | NA19171    | Yoruba    | *2/*41 | *2/*41                                   | *2/*41          | —      | 3           | *2x2/*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/*3  | *1/*3                                    | *1/*3             | —      | 2           | *1/*3                    |
| 45        | NA19175    | Yoruba    | *1/*4  | *1/*4                                    | *1/*4             | 4      | 1           | *1/*4x3                  |
| 46        | NA19147    | Yoruba    | *17/*29| *17/*29                                  | *17/*29          | 2      | *17/*29                 |
| 47        | NA18505    | Yoruba    | *1/*45 | *1/*45                                   | *1/*45           | 2      | *1/*45                   |
| 48        | NA18517    | Yoruba    | *10/*10| *10/*10                                  | *10/*10         | 2      | *10/*10                  |

Abbreviations: CNV, copy number variation; LR-PCR, long-range PCR; RETINA, PCR-based real-time invader assay.
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 do not detect 31 G\rightarrow\text{A}. Roche Amplichip had the best allele coverage among the platforms used in this previous study; however, it does not include the defining CYP2D6*41 SNP 2988G\rightarrow\text{A} and thus may misclassify some CYP2D6*2 alleles as CYP2D6*41. Additionally, in that study, discrepant genotypes (which did exist between different methods) were not adjudicated/confirmed by Sanger or other sequencing methods. Our study, in contrast, has several significant advantages. First, one of our genotyping methods—RETINA—covers 29 of the most frequent 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 US 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 10 additional loci (124 G\rightarrow\text{A}[\text{CYP2D6*12}], 4142_4133\,\text{dupGTGCCCACT}[\text{CYP2D6*18}], 2573_2574\,\text{insC}[\text{CYP2D6*21}], 2587_2590\,\text{delGACT}[\text{CYP2D6*38}], 77\,\text{G}\rightarrow\text{A}[\text{CYP2D6*43}], 1716\,\text{G}\rightarrow\text{A}[\text{CYP2D6*45}], 77\,\text{G}\rightarrow\text{A}\,\text{and}\,1716\,\text{G}\rightarrow\text{A}[\text{CYP2D6*46}], 3259_3260\,\text{insGT}[\text{CYP2D6*42}],

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            | NA        | AA        | *4/*4  | —      | 3            | *4x2/*4                  | 0                        | PM                        |
| 2L            | UC9208    | AA        | *1/*29 | —      | 2            | *1/*29                   | 1.5                      | EM                        |
| 3L            | NA        | Caucasian | *4/*41 | —      | 2            | *4/*41                   | 0.5                      | IM                        |
| 4L            | NA        | Caucasian | *4/*41 | —      | 2            | *4/*41                   | 0.5                      | IM                        |
| 5L            | NA        | Caucasian | *2/*41 | —      | 2            | *2/*41                   | 1.5                      | EM                        |
| 6L            | UC9305    | Caucasian | *1/*4  | —      | 2            | *1/*4                    | 1.0                      | EM                        |
| 7L            | NA        | Caucasian | *1/*6  | —      | 2            | *1/*6                    | 1.0                      | EM                        |
| 8L            | NA        | Caucasian | *1/*4  | —      | 3            | *1/*4x2                  | 1.0                      | EM                        |
| 9L            | NA        | AA        | *4/*17 | —      | 3            | *4x2/*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            | *2x2/*41                 | 2.5                      | UM                        |
| 14L           | NA        | 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  | *S     | 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                        |
| 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  | *S     | 2            | *1x2/*S                 | 2.0                      | EM                        |
| 33L           | HH840     | Caucasian | *4/*4  | —      | 2            | *4/*4                    | 0                        | PM                        |
| 34L           | HH841     | Caucasian | *2/*2  | *S     | 1            | *2/*S                    | 1.0                      | EM                        |
| 35L           | HH844     | Caucasian | *4/*1  | —      | 3            | *1/*4x2                  | 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| *S     | 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           | NA        | Caucasian | *4/*4  | —      | 2            | *4/*4                    | 0                        | PM                        |
| 44L           | NA        | Caucasian | *4/*4  | —      | 2            | *4/*4                    | 0                        | PM                        |
| 45L           | NA        | Caucasian | *35/*41| —      | 2            | *35/*41                  | 1.5                      | EM                        |
| 46L           | NA        | Caucasian | *1/*1  | *S     | 1            | *1/*5                    | 1.0                      | EM                        |
| 47L           | NA        | Caucasian | *1/*1  | *S     | 1            | *1/*1                    | 1.0                      | EM                        |
| 48L           | NA        | Caucasian | *1/*4  | —      | 2            | *1/*4                    | 1.0                      | EM                        |
| 49L           | NA        | Caucasian | *4/*1  | —      | 2            | *4/*1                    | 0.5                      | IM                        |
| 50L           | NA        | Caucasian | *6/*35 | —      | 2            | *6/*35                   | 1.0                      | EM                        |

Abbreviations: AA, African American; CNV, copy number variation; EM, extensive metabolizer; IM, intermediate metabolizer; LR-PCR, long-range PCR; NA, not available; PM, poor metabolizer; RETINA, PCR-based real-time invader assay; UM, ultrarapid metabolizer.
Second, we used the RETINA assays for which positive signals were not observed in our samples that had unexpectedly low metabolizing activity compared with assigned AS 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, $P = 4.2 \times 10^{-8}$; $r_s = 0.68$ by rank correlation test, $P = 5.3 \times 10^{-4}$). The quantitation limit was $0.4 \text{ pmol min}^{-1} \text{ng}^{-1}$ protein for endoxifen formation rate.

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 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.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**ACKNOWLEDGMENTS**

We acknowledge the help of Dr R Stephanie Huang and Bonnie LaCroix from the Pharmacogenomics of Anticancer Agents Cell Core and the kind remarks from Dr Kazuma Kayotani. We also thank the Liver Tissue Procurement and Distribution System (NIH contract 3N01-DK-9-2310) and the Cooperative Human Tissue Network for providing the liver samples. This work is supported by NIH U01GM061393 (Pharmacogenomics of Anticancer Agents Research Group; to YN, MTR and NJC) and NIH K12 CA139160 and K23 GM100288-01A1 (to PHO). HF is supported by a NIH T32 GM07228-01A1 grant from the National Institute of General Medical Sciences. FL is supported by a NIH T32 GM07228-01A1 grant from the National Institute of General Medical Sciences. YN serves as the Director of the Patient Outcomes Research Team and a recipient of a Conquer Cancer Foundation of ASCO Translational Research Professorship, in memory of Merrill J Egorin, MD. Any opinions, findings and conclusions expressed in this material are those of the authors and do not necessarily reflect those of the American Society of Clinical Oncology or the Conquer Cancer Foundation.

**REFERENCES**

1 Zhou SF. Polymorphism of human cytochrome P450 2D6 and its clinical significance: part II. Clin Pharmacokinet 2009; 48: 761-804.
