Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Pharmacogenomics

Integrated *CYP2D6* interrogation for multiethnic copy number and tandem allele detection

Wanqiong Qiao[1,2], Suparna Martis[1], Geetu Mendiratta[1,2], Lisong Shi[1,2], Mariana R Botton[1,2], Yao Yang[1], Andrea Gaedigk[3], Raymon Vijzelaar[4], Lisa Edelmann[1,2], Ruth Kornreich[1,2], Robert J Desnick[1] & Stuart A Scott*[,1,2]

[1]Department of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[2]Sema4, a Mount Sinai venture, Stamford, CT 06902, USA

[3]Division of Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy Kansas City, Kansas City, MO 64108, USA

[4]MRC-Holland, Willem Schoutenstraat 6, Amsterdam, The Netherlands

*Author for correspondence: Tel.: +1 212 241 3780; Fax: +1 212 241 1464; stuart.scott@mssm.edu

Aim: To comprehensively interrogate *CYP2D6* by integrating genotyping, copy number analysis and novel strategies to identify *CYP2D6*36* and characterize *CYP2D6* duplications. **Methods:** Genotyping of 16 *CYP2D6* alleles, multiplex ligation-dependent probe amplification (MLPA) and *CYP2D6*36* and duplication allele-specific genotyping were performed on 427 African–American, Asian, Caucasian, Hispanic, and Ashkenazi Jewish individuals. **Results:** A novel PCR strategy determined that almost half of all *CYP2D6*10* (100C>T) alleles are actually **36* (isolated or in tandem with **10*) and all identified duplication alleles were characterized. Integrated results from all testing platforms enabled the refinement of genotype frequencies across all studied populations. **Conclusion:** The polymorphic *CYP2D6* gene requires comprehensive interrogation to characterize allelic variation across ethnicities, which was enabled in this study by integrating multiplexed genotyping, MLPA copy number analysis, novel PCR strategies and duplication allele-specific genotyping.

First draft submitted: 19 August 2018; Accepted for publication: 8 October 2018; Published online: 6 December 2018

Keywords: copy number variation • *CYP2D6* • *CYP2D6*36* • genotyping • MLPA • pharmacogenetics

The CYP2D6 enzyme is involved in the hepatic metabolism of approximately 25% of commonly prescribed drugs, including antidepressants, antipsychotics, antiarrhythmics, opioids and β-blockers [1]. Variant *CYP2D6* star (*) alleles encode enzymes with increased, decreased or no activity, which can contribute to interindividual variability in drug response and adverse reaction risk. Multiple techniques have been employed to interrogate *CYP2D6* genotype and copy number, including long-range PCR and targeted genotyping, allele-specific PCR, quantitative real-time PCR, Sanger sequencing and long-read single molecule real-time (SMRT) sequencing [2–5]. The *CYP2D6* gene is highly polymorphic with over a hundred star (*) alleles currently cataloged by the Pharmacogene Variation (PharmVar) Consortium (https://www.pharmvar.org/gene/CYP2D6) [6], which include single nucleotide variants, small insertions/deletions, gene conversions, copy number variants (e.g., *CYP2D6*5, *36×2*) and 'tandem' allele structural rearrangements (e.g., *CYP2D6*36+*10*). However, the highly homologous neighboring *CYP2D7* and *CYP2D8* pseudogenes make interrogating *CYP2D6* technically challenging, which ultimately can lead to inconsistencies in results between genotyping platforms and potential inaccuracies in CYP2D6 metabolizer phenotype prediction [7,8].

The decreased function *CYP2D6*10* haplotype is prevalent in populations of Asian ancestry; however, it can also be indicative of an undetected **36* or tandem **36+*10* allele due to shared variants that are commonly included in *CYP2D6* genotyping platforms (i.e., 100C>T; 1661G>C; 4180G>C) [9–11]. The *CYP2D6*36* allele differs from **10* by the presence of an exon 9 conversion that is not typically interrogated by common genotyping platforms [9]. The *CYP2D6*10* allele is included in several genotyping platforms (Affymetrix DMET, Luminex

xTAG, LifeTech TaqMan laboratory-developed assay, Agena Bioscience iPLEX ADME PGx Pro, Agena Bioscience CYP2D6 and Autogenomics CYP2D6); however, of these assays only the Agena Bioscience *CYP2D6* panel directly interrogates *CYP2D6*36* [9]. A PCR-based assay to detect *CYP2D6*36* and **36* duplications (**36×N*) has been reported [10]; however, this assay cannot identify an exon 9 conversion in the context of a tandem **36+*10* allele. Long-range PCR can also be utilized to infer the presence of *CYP2D6*36+*10*, as the tandem allele produces an approximately 10 kb long 'fragment D' encompassing the **36* gene copy and a *CYP2D7*-derived exon 9 and downstream region that make this amplicon 1.7 kb longer than a 'fragment D' generated from a 'typical' duplication (e.g., *CYP2D6*2×2* or a **10×2*) [12].

Another challenge with *CYP2D6* genotyping is specifically interrogating duplicated alleles, as accurate prediction of CYP2D6 metabolizer status necessitates direct analysis of the gene copy (or copies) when an increased copy number is detected, particularly when identified concurrently with normal activity and no function alleles in compound heterozygosity (e.g., *CYP2D6*1/*4, Dup*) [2,13]. The *CYP2D6* alleles that have been reported as duplications are listed within the structural variation document on the PharmVar website (https://www.pharmvar.org/gene/CYP2D6), and their activity scores are determined by multiplying the score of the specific star (*) allele by the number of copies (e.g., activity score of **41×2* is one [0.5 × 2]) [14]. The activity score system assigns values to specific alleles as a reflection of their activity, and the sum of the values of both alleles provides the activity score of a diplotype, which ultimately facilitates the translation of *CYP2D6* diplotype to phenotype [15].

To address these technical challenges, we developed a rapid PCR strategy that can distinguish *CYP2D6*36* and/or **36+*10* alleles from the related **10*, as well as a duplication allele-specific genotyping application of a commercially available *CYP2D6* genotyping assay. These were validated by orthogonal Sanger sequencing and tested across a large multiethnic cohort from the New York City metropolitan area (USA).

Materials & methods

Subjects

Peripheral blood samples from healthy adult donors who self-reported their racial and ethnic background (African–American [AA], Asian, Caucasian or Hispanic) and gave informed consent for the use of their DNA for research were obtained from the New York Blood Center (NY, USA) with Institutional Review Board approval as previously described [16]. In addition, blood samples were obtained with informed consent from unrelated, healthy Ashkenazi Jewish (AJ) individuals from the greater New York City metropolitan area as previously described [17–19]. All personal identifiers were removed, and isolated DNA samples were tested anonymously. Genomic DNA was isolated using the Puregene® DNA Purification kit (Qiagen, CA, USA) according to the manufacturer's instructions.

Genotyping

The *CYP2D6* allele designations refer to those defined by the PharmVar Consortium (https://www.pharmvar.org). Genotyping of 16 variant *CYP2D6* alleles (**2 - *11, *14, *15, *17, *29, *35, *41*) and the gene duplication was performed using the xTAG CYP2D6 Kit v3 (Luminex Corporation, TX, USA) according to the manufacturer's instructions and as previously described [20]. Notably, the wild-type **1* allele was assigned in the absence of other detectable variant alleles.

Copy number analysis

CYP2D6 copy number at exons 1, 4, 6 and the 3′ downstream region was interrogated by multiplex ligation-dependent probe amplification (MLPA) using the SALSA MLPA P128-B1 Cytochrome P450 kit (MRC-Holland, Amsterdam, The Netherlands) according to the manufacturer's instructions and as previously described [21]. Of note, interrogating multiple loci across the *CYP2D6* gene is necessary to detect copy number signatures that are indicative of structurally rearranged and tandem alleles (e.g., **36+10, *68+*4*, etc.), in addition to the standard full gene deletion and duplication alleles. Raw data was analyzed using GeneMarker v1.90 software (SoftGenetics, PA, USA) and copy number was determined based on the following criteria: peak ratio $\geq$0.25 and $<$0.75 for one copy, $\geq$0.75 and $<$1.25 for two copies, $\geq$1.25 and $<$1.7 for three copies, $\geq$1.7 and $<$2.2 for four copies and $\geq$2.2 for more than four copies.

Distinguishing *CYP2D6*36* alleles from **10*

To distinguish the *CYP2D6*36* exon 9 conversion among the samples that genotyped as the related **10* (i.e., 100C>T; 1661G>C; 4180G>C), a novel PCR strategy was developed (Figure 1). A set of primers were

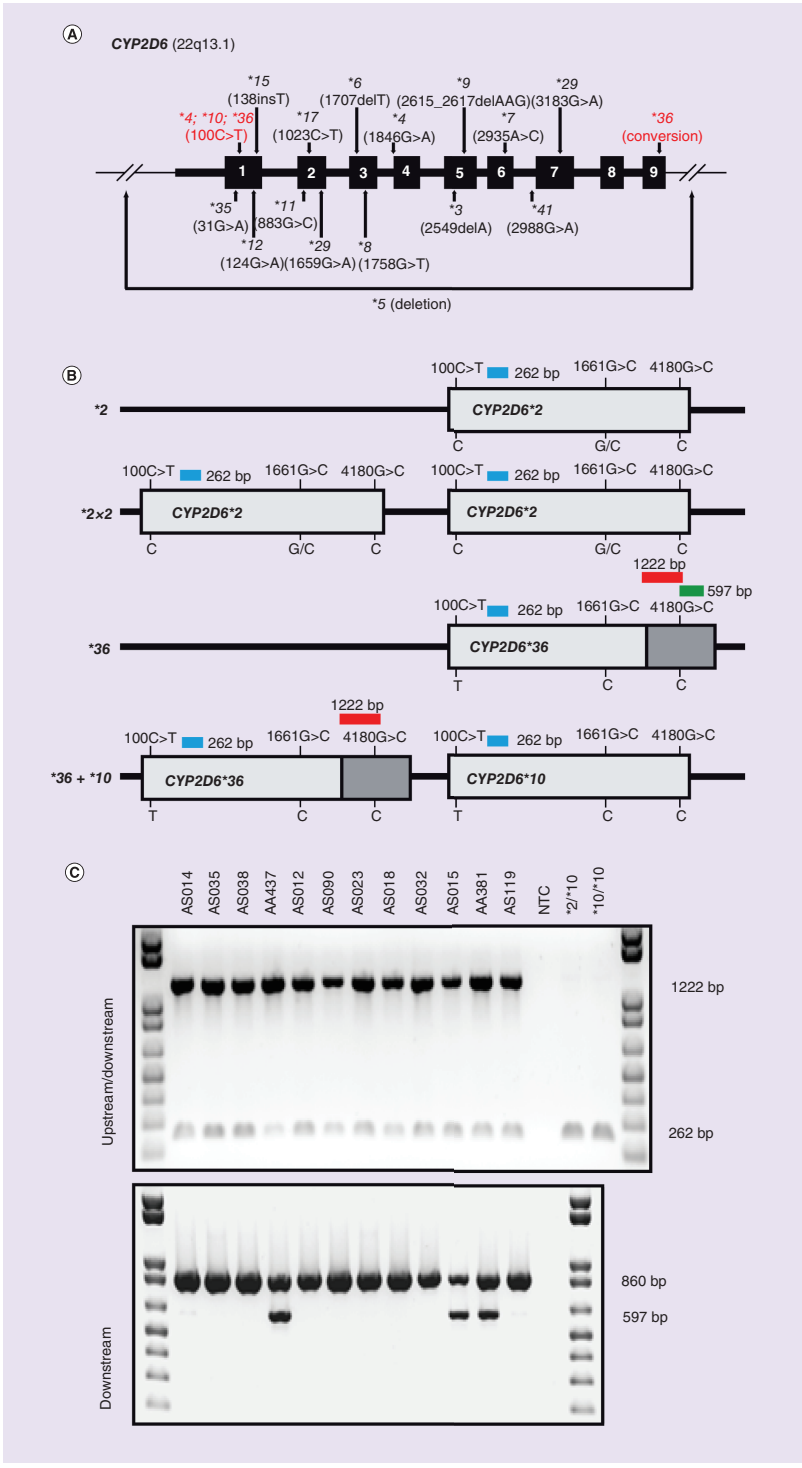

Figure 1. *CYP2D6* gene structure, targeted genotyping, and **36* allele identification. (A) Gene diagram of *CYP2D6* highlighting the location of variant star (*) alleles that are commonly included in targeted genotyping assays, including the deletion allele (**5*), 100C>T and the exon 9 conversion. Variant positions are denoted by their common nucleotide nomenclature from M33388.1 GenBank reference sequence. **(B)** Illustration of the *CYP2D6*36* PCR assay highlighting the ability to distinguish **36* and **36+*10* among individuals who were genotyped as **10* carriers (i.e., 100C>T; 1661G>C; 4180G>C). Dark gray bars represent the 3′ downstream conversion regions **(C)** (Top panel) Upstream/downstream duplex PCR results indicating the presence of an exon 9 conversion in all 12 tested samples (1222 bp product; red bar in the bottom panel). All samples also amplified the 262 bp *CYP2D6* control product, including the **36*-negative controls. (Bottom panel) Downstream duplex PCR results indicating the presence of a downstream exon 9 conversion in samples AA437, AS015 and AA381 (597 bp product; green bar in panel B). All samples amplified the 860 bp *CYP3A7* control fragment. The combined upstream/downstream and downstream PCR results enable the detection of *CYP2D6*36* and **36+*10* carriers.
bp: Base pair; NTC: No template control.
For color figures please see online at: https://www.futuremedicine.com/doi/full/10.2217/pgs-2018-0135

Table 1. Primers and assay conditions for *CYP2D6* PCR analysis.

Primer sequence (5′–3′)	Amplicon size		Annealing temperature	Ref.
	CYP2D6 exon 9	Exon 9 conversion		
Upstream/downstream exon 9 conversion primers				
GTGAGCCCATCTGGGAAACA				
AAGCTGACGACACGAGAGTG	–	1222 bp		
GATGACCGTAGTCCGAGCTG	262 bp	262 bp	62°C	
CCCACACTGAGCTTACAGCA				
Downstream exon 9 conversion primers				
AGCCACTCTCGTGTCGTCAGCTT				[10]
CGACTGAGCCCTGGGAGGTAGGTAG	–	597 bp		[10]
CACCTCTGCTAAGGGAAACAGGCC	860 bp	860 bp	68°C	[10]
GCCAGCCTGAACATCCTTTTTGCTA				[10]
Upstream/downstream copy primers				
GAACCTCTGGAGCAGCCCATACCC				[2]
ACTGAGCCCTGGGAGGTAGGTAG	~5 kb		68°C	[12]
Downstream copy primers ('Fragment D' Fwd; 'Fragment F' Rev)				
CCAGAAGGCTTTGCAGGCTTCAG				[12]
CAGGCATGAGCTAAGGCACCCAGA	~8.1 kb		63°C	[12]
Upstream copy primers ('Fragment D')				
CCAGAAGGCTTTGCAGGCTTCAG				[12]
CGGCAGTGGTCAGCTAATGAC	~8.6 kb or 10.2 kb		63°C	[12]
Exon 8/9 Sanger sequencing primers				
CCGTCTAGTGGGGAGACAAA				
GTGGGGTAAGCAGGAATGAG	842 bp		60°C	

bp: Base pair; kb: Kilobase.

designed that amplify a 1222 base pair (bp) product when an exon 9 conversion occurs in either an 'upstream' or 'downstream' *CYP2D6* copy (Table 1). This reaction was duplexed with a *CYP2D6* internal control 262 bp amplicon. In addition, previously reported primers were used to concurrently amplify a 597 bp product when an exon 9 conversion occurs only in a 'downstream' *CYP2D6* copy, which was duplexed with an internal control 860 bp amplicon [10]. The combined results of these two PCR assays can therefore identify and determine if an exon 9 conversion is located in an 'upstream' or 'downstream' *CYP2D6* copy.

To validate the *CYP2D6*36* PCR assay, long-range PCR of both 'downstream' and 'upstream' *CYP2D6* gene copies (when present) were performed and amplicons subjected to Sanger sequencing. Long-range PCR of the 'downstream' copy was performed in 20 µl containing approximately 40 ng of DNA, 1× SequalPrep™ Reaction buffer (Invitrogen, CA, USA), 0.5 µM of forward and reverse primers (Table 1) and 1.8 units of SequalPrep™ Polymerase. Amplification consisted of an initial denaturation step at 94°C for 2 min followed by ten amplification cycles (94°C for 10 s, 63°C for 30 s and 68°C for 13 min), another 20 amplification cycles (94°C for 10 s, 63°C for 30 s and 68°C for 13 min + 20 s/cycle) and a final extension at 72°C for 5 min. A second long-range PCR that amplified all gene copies, that is, one or more duplicated 'upstream' gene copies as well as the 'downstream' gene copy was also employed. This reaction amplified a 5.0 kb fragment as above but with annealing and extension at 68°C, and an extension time of 6 min given the shorter amplicon length. Amplicons were subjected to Sanger sequencing with primers targeting *CYP2D6* exon 9, and results were analyzed using FinchTV (Geospiza, WA, USA).

CYP2D6 duplication allele-specific genotyping

Duplicated copies of *CYP2D6* were directly interrogated by the long-range PCR that specifically amplified 'upstream' copies of the gene [2,12]. As previously described, these primers specifically amplified an 8.6 kb fragment (also known as 'fragment D' [12]) that encompasses the entire 'upstream' *CYP2D6* copy (or 10.2 kb in the presence of the **36* allele), allowing for star (*) allele determination of duplicated copies when present [2]. The PCR conditions for the 'upstream' long-range PCR were identical to the 'downstream' copy amplification detailed above. Both 'upstream'

Table 2. Representative *CYP2D6*10* cases that were revised by exon 9 conversion PCR and MLPA analysis.

Original diplotype	Exon 9 conversion PCR		MLPA (copy number[†])				Revised diplotype
Upstream/ downstream		Downstream	Exon 1	Exon 4	Exon 6	Downstream	
**1/*10*	+		3	3	3	2	**1/*36+*10*
**1/*10*	+		4	4	4	2	**1/*36×2+*10*
**2/*10*	+		3	3	3	2	**2/*36+*10*
**2/*10*	+	+	2	2	2	2 [‡]	**2/*36*
**4/*10*	+		3	3	3	2	**4/*36+*10*
**5/*10*	+		2	2	2	1	**5/*36+*10*
**5/*10*	+		3	3	3	1	**5/*36×2+*10*
**10/*10*	+		3	3	3	2	**10/*36+*10*
**10/*10*	+		4+	4+	4+	2	**10/*36×N+*10*
**10/*10*	+	+	4	4	4	2	**36/*36×2+*10*
**10/*17*	+	+	2	2	2	2 [‡]	**17/*36*
**10/*41*	+		3	3	3	2	**41/*36+*10*

[†] '4+' represents a total copy number of at least four by MLPA.
[‡] Representative *CYP2D6*36* sample with normal 'downstream' region copy number by MLPA testing (see Discussion).
MLPA: Multiplex ligation-dependent probe amplification.

and 'downstream' amplicons were used as templates for a 5.0 kb nested PCR as detailed above, which subsequently were subjected to shrimp alkaline phosphatase (SAP)/exonuclease digestion and multiplexed allele-specific primer extension genotyping using the allele-specific primer extension and bead hybridization protocol of the xTAG CYP2D6 Kit v3 (Luminex Molecular Diagnostics, Toronto, ON, Canada).

Results

CYP2D6 genotyping & copy number analysis in a multiethnic cohort

Multiplexed targeted genotyping and MLPA analysis of 427 AA (n = 80), Asian (n = 87), Caucasian (n = 83), Hispanic (n = 81) and AJ (n = 96) individuals identified the frequencies of 16 variant *CYP2D6* alleles and the copy number of exons 1, 4, 6 and 3′ downstream region (Supplementary Table 1). However, the absence of any genotyping probes for the *CYP2D6*36* exon 9 conversion and the inability to definitively discriminate duplicated alleles by targeted genotyping or MLPA prompted the development of additional assays to further refine *CYP2D6* allele frequencies in our multiethnic cohort.

Detection of *CYP2D6*36* alleles

To distinguish the *CYP2D6*36* allele series from the identified **10* carriers (100C>T; 1661G>C; 4180G>C), our PCR strategy was applied to all subjects that carried **10* based on Luminex genotyping. As illustrated in Figure 1, in the presence of an exon 9 conversion, both 1222 and 262 bp fragments amplified using the upstream/downstream *CYP2D6*36* primer pair (Figure 1A). If no upstream or downstream exon 9 conversion was present, only the 262 bp control fragment amplified. Similarly, if the exon 9 conversion occurs in a downstream copy, both the 597 bp fragment and the 860 bp control fragment amplified using the downstream-specific *CYP2D6*36* primers (Figure 1B). If no downstream exon 9 conversion was present, only the 860 bp control fragment amplified. The integrated results of this PCR strategy refined 43 of 94 (45.7%) detected **10* alleles into **36* (n = 3; 3.2%), **36+*10* (n = 30; 31.9%) and **36×N+*10* (n = 10; 10.6%), the majority of which being derived from the Asian cohort.

These results were further supported by MLPA, which indicated a 3′ downstream copy number loss across most carriers (see Discussion) as well as copy number gains at exons 1, 4 and 6 among the tandem allele carriers (Table 2). The results of the *CYP2D6*36* PCR assay were also confirmed by long-range PCR and bidirectional Sanger sequencing of both upstream and downstream copies on 24 randomly selected Asian samples that were initially genotyped as **10* (data not shown).

Table 3. *CYP2D6* duplication allele-specific genotyping.

Original diplotype	Copy-specific genotyping		MLPA copy number	Revised diplotype	Activity score
	Upstream	Downstream			
**1/*2, Dup*	**1*	**1/*2*	3	**1×2/*2*	3
**4/*17, Dup*	**4*	**4/*17*	3	**4×2/*17*	0.5
**2/*4, Dup*	**4*	**2/*4*	3	**2/*4×2*	1
**2/*4, Dup*	**2*	**2/*4*	3	**2×2/*4*	2
**2/*4, Dup*	**4*	**2/*4*	3	**2/*4×2*	1
**1/*4, Dup*	**4*	**1/*4*	3	**1/*4×2*	1
**1/*4, Dup*	**4*	**1/*4*	3	**1/*4×2*	1
**4/*17, Dup*	**4*	**4/*17*	3	**4×2/*17*	0.5
**1/*2, Dup*	**2*	**1/*2*	3	**1/*2×2*	3
**2/*41, Dup*	**2*	**2/*41*	3	**2×2/*41*	2.5
**1/*2, Dup*	**2*	**1/*2*	3	**1/*2×2*	3
**1/*2, Dup*	**2*	**1/*2*	4	**1/*2×3*	4
**2/*9, Dup*	**2*	**2/*9*	3	**2×2/*9*	2.5
**1/*2, Dup*	**1/*2*	**1/*2*	4	**1×2/*2×2*	4
**1/*2, Dup*	**2*	**1/*2*	3	**1/*2×2*	3
**1/*4, Dup*	**1*	**1/*4*	3	**1×2/*4*	2
**2/*41, Dup*	**2*	**2/*41*	3	**2×2/*41*	2.5
**1/*2, Dup*	**2*	**1/*2*	3	**1/*2×2*	3
**2/*4, Dup*	**2*	**2/*4*	3	**2×2/*4*	2
**1/*2, Dup*	**2*	**1/*2*	3	**1/*2×2*	3
**2/*35, Dup*	**2*	**2/*35*	3	**2×2/*35*	3
**1/*4, Dup*	**1*	**1/*4*	3	**1×2/*4*	2
**1/*2, Dup*	**2*	**1/*2*	3	**1/*2×2*	3
**2/*17, Dup*	**2*	**2/*17*	3	**2×2/*17*	2.5
**1/*4, Dup*	**1*	**1/*4*	3	**1×2/*4*	2
**2/*41, Dup*	**2*	**2/*41*	3	**2×2/*41*	2.5

CYP2D6 duplication allele-specific genotyping

All samples with *CYP2D6* copy number gains by Luminex and MLPA (with the exception of **10* allele carriers) were further subjected to duplication allele-specific multiplexed targeted genotyping (n = 26). Upstream and downstream copies were amplified for each sample and both amplicons were used independently as templates for Luminex genotyping. This strategy unambiguously identified the haplotype of each duplicated *CYP2D6* copy among all tested samples, including 16 samples (61.5%) that were compound heterozygous for *CYP2D6* alleles with different activity scores (e.g., *CYP2D6*2/*4, Dup*; Table 3).

Revised *CYP2D6* allele & genotype frequencies

The results from the *CYP2D6*36* exon 9 and duplication allele-specific genotyping were incorporated into the previously identified multiethnic *CYP2D6* allele and genotype frequencies, which are summarized in Table 4 and Supplementary Table 2, respectively. The *CYP2D6*7, *8, *11* and **15* alleles were not detected in any of the tested subjects. The combined frequencies of detected alleles in the AA, Asian, Caucasian, Hispanic and AJ populations were distributed as increased function (0.038, 0.011, 0.012, 0.024 and 0.104), normal function (0.475, 0.426, 0.596, 0.725 and 0.438), decreased function (0.300, 0.432, 0.175, 0.116 and 0.240) and no function (0.188, 0.120, 0.217, 0.128 and 0.219), respectively (Table 4 & Figure 2A).

Prediction of CYP2D6 phenotype from genotype data

There currently is no standardized protocol for the translation of *CYP2D6* genotype to phenotype, which mostly affects diplotypes consisting of one functional and one no function, or two decreased function alleles (i.e., total activity score = 1). These diplotypes are often classified in the literature as either normal metabolizers (NMs) or intermediate metabolizers, which recently has been highlighted by the Clinical Pharmacogenetics Implementation

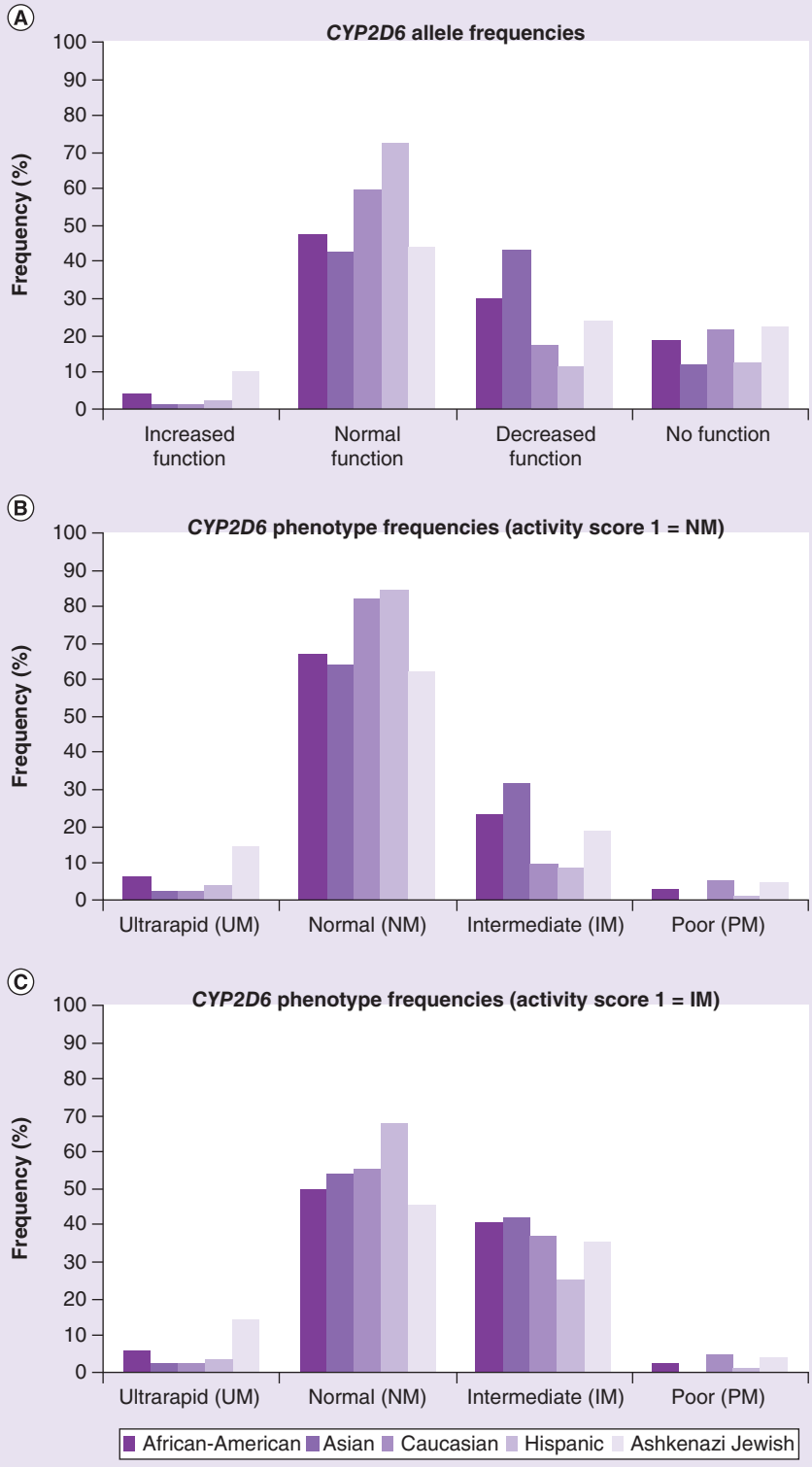

Figure 2. Graphical illustration of the identified multiethnic *CYP2D6* allele frequencies (A), and metabolizer phenotype frequencies using two systems, activity score of 1 = normal metabolizers (B) and activity score of 1 = intermediate metabolizers (C).

Table 4. Revised *CYP2D6* allele frequencies.

CYP2D6 allele	African–American (n = 160)		Asian (n = 174)		Caucasian (n = 166)		Hispanic (n = 162)		Ashkenazi Jewish (n = 192)	
	Freq.	95% CI	Freq.	95% CI	Freq.	95% CI	Freq.	95% CI	Freq.	95% CI
Increased function										
*1×2	0.013	0.000–0.030	0.000	0.000–0.000	0.000	0.000–0.000	0.006	0.000–0.018	0.026	0.004–0.049
*2×2	0.025	0.001–0.049	0.011	0.000–0.027	0.012	0.000–0.029	0.018	0.000–0.039	0.078	0.040–0.116
Subtotal	0.038		0.011		0.012		0.024		0.104	
Normal function										
*1	0.256	0.189–0.324	0.273	0.207–0.339	0.361	0.288–0.435	0.475	0.398–0.552	0.313	0.247–0.378
*2	0.213	0.149–0.276	0.142	0.090–0.194	0.169	0.112–0.226	0.226	0.162–0.290	0.099	0.057–0.141
*29×2	0.006	0.000–0.018	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000
*35	0.000	0.000–0.000	0.011	0.000–0.027	0.066	0.028–0.104	0.024	0.001–0.048	0.026	0.004–0.049
Subtotal	0.475		0.426		0.596		0.725		0.438	
Decreased function										
*9	0.000	0.000–0.000	0.006	0.000–0.017	0.012	0.000–0.029	0.024	0.001–0.048	0.005	0.000–0.015
*10	0.044	0.012–0.075	0.136	0.086–0.187	0.030	0.004–0.056	0.000	0.000–0.000	0.078	0.040–0.116
*17	0.125	0.074–0.176	0.000	0.000–0.000	0.000	0.000–0.000	0.006	0.000–0.018	0.010	0.000–0.025
*29	0.106	0.059–0.154	0.006	0.000–0.017	0.000	0.000–0.000	0.006	0.000–0.018	0.000	0.000–0.000
*36+*10	0.000	0.000–0.000	0.153	0.100–0.207	0.006	0.000–0.018	0.006	0.000–0.018	0.005	0.000–0.015
*36×2+*10	0.000	0.000–0.000	0.057	0.023–0.092	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000
*41	0.025	0.001–0.049	0.074	0.035–0.113	0.127	0.076–0.177	0.073	0.033–0.113	0.141	0.091–0.190
*41×2	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000
Subtotal	0.300		0.432		0.175		0.116		0.240	
No function										
*3	0.000	0.000–0.000	0.006	0.000–0.017	0.018	0.000–0.038	0.006	0.000–0.018	0.005	0.000–0.015
*4	0.069	0.030–0.108	0.051	0.019–0.084	0.169	0.112–0.226	0.091	0.047–0.136	0.203	0.146–0.260
*4×2	0.050	0.016–0.084	0.000	0.000–0.000	0.006	0.000–0.018	0.000	0.000–0.000	0.000	0.000–0.000
*5	0.056	0.021–0.092	0.046	0.015–0.077	0.018	0.000–0.038	0.030	0.004–0.057	0.005	0.000–0.015
*6	0.000	0.000–0.000	0.000	0.000–0.000	0.006	0.000–0.018	0.000	0.000–0.000	0.005	0.000–0.015
*7	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000
*8	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000
*11	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000
*14	0.000	0.000–0.000	0.011	0.000–0.027	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000
*15	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000
*36	0.013	0.000–0.030	0.006	0.000–0.017	0.000	0.000–0.000	0.000	0.000–0.000	0.000	0.000–0.000
Subtotal	0.188		0.120		0.217		0.128		0.219	

Freq.: Frequency.

Consortium (CPIC) guideline for *CYP2D6* and tamoxifen [22]. Classifying patients with an activity score of 1 as NMs results in predicted *CYP2D6* metabolizer phenotypes in the AA, Asian, Caucasian, Hispanic and AJ of ultrarapid (6.3, 2.3, 2.4, 3.7 and 14.6%), normal (67.5, 64.8, 83.1, 85.4 and 62.5%), intermediate (23.8, 31.8, 9.6, 8.5 and 18.8%) and poor (2.5, 0.0, 4.8, 1.2 and 4.2%) metabolizers, respectively (Supplementary Table 2 &Figure 2B). However, classifying patients with an activity score of 1 as intermediate metabolizers results in modified *CYP2D6* metabolizer phenotypes in the AA, Asian, Caucasian, Hispanic and AJ of normal (50.0, 54.5, 55.4, 68.3 and 45.8%) and intermediate (41.3, 42.1, 37.3, 25.6 and 35.4%) metabolizers, respectively (Figure 2C).

Discussion

The technical challenges with interrogating the polymorphic *CYP2D6* gene prompted our development of a novel PCR assay for the exon 9 conversion that unequivocally allows us to detect the *CYP2D6*36* allele, as well as a genotyping strategy that specifically characterizes *CYP2D6* duplication alleles. In conjunction with multiplexed *CYP2D6*-targeted genotyping and MLPA analysis, this integrated testing approach was performed on a large cohort of AA (n = 80), Asian (n = 87), Caucasian (n = 83), Hispanic (n = 81) and AJ (n = 96) individuals from the New

York City metropolitan area. Despite the polymorphic nature and structural complexity of the *CYP2D6* gene, our comprehensive genotyping and copy number interrogation (requiring ~400–500 ng of DNA) enabled refined allele and diplotype detection across all tested populations. Moreover, the inclusion of our exon 9 conversion PCR assay determined that almost half of all *CYP2D6*10* alleles are actually unrecognized **36* or **36* tandem alleles, which are not directly interrogated by most commercial genotyping panels.

Despite the increasing availability of high-throughput short-read pharmacogenomic sequencing programs [23,24], the challenges with accurately interrogating the *CYP2D6* locus and the uncertainty of interpreting both novel and rare variants [3,25] together suggest that targeted genotyping of functional *CYP2D6* variants will continue to be used by both clinical and research laboratories. Similarly, the recent development of long-read *CYP2D6* SMRT sequencing using the Pacific Biosciences platform [2] is currently only accessible to those with the necessary sequencing infrastructure and bioinformatics expertise. As such, at the center of this study was the targeted genotyping of 16 well-characterized variant *CYP2D6* alleles (including gene deletion and duplication), coupled with quantitative copy number interrogation at four loci across the *CYP2D6* gene by MLPA. These multiplexed commercial assays are robust and accurate, and their results complement each other; however, their inability to directly genotype the exon 9 conversion and identify the sequence of duplicated *CYP2D6* alleles prompted our development of alternative strategies for more comprehensive characterization.

The 100C>T (p.P34S) transition was one of the first variant *CYP2D6* alleles reported [26–28] and later cataloged among several human *CYP2D6* haplotypes [6,29]. This common variant is located in exon 1 of the *CYP2D6* gene (NM_000106.5:c.100C>T; rs1065852), and has a global minor allele frequency (MAF) of 20.5% in the Genome Aggregation Database (http://gnomad.broadinstitute.org/), which is highest among east Asians with a MAF of 57.8% (European: 24.9%; south Asian: 18.1%; African: 15.1% and Latino: 15.1%) [30]. When 100C>T is in *cis* with only common 'benign' *CYP2D6* variants (i.e., 1661G>C,4180G>C), the haplotype is defined as *CYP2D6*10* and classified as decreased function. Of note, an international working group has recently suggested revising the activity score of *CYP2D6*10* from 0.5 to 0.25 (https://cpicpgx.org/resources/cyp2d6-genotype-to-phenotype-standardization-project/).

However, 100C>T (p.P34S) is also an important variant of the nonfunctional **4* haplotype (with 1846G>A), as well as several other nonfunctional (**14, *36, *47, *56, *57, *69, *99, *100, *101*), decreased (**49, *54, *72*) and uncertain (**37, *52, *65, *87, *94, *95*) function haplotypes. The prevalence and functionality of 100C>T prompted its inclusion into most commercial *CYP2D6* genotyping assays and laboratory-developed tests; however, it is not typically differentiated by these platforms beyond the *CYP2D6*10* and **4* haplotypes. Although this is most likely due to the low or unknown population frequencies of the other star (*) allele variants that also harbor 100C>T [31], notable among them is the nonfunctional *CYP2D6*36* that is principally defined by 100C>T and an exon 9 conversion [10]. This related allele (often found as a tandem **36+*10*) is more prevalent among east Asians (~27%) [31,32], but has also been found in Africans and AAs (~1%) [31]; however, it is rarely included in commercial genotyping platforms.

Our multiethnic population screen detected 100C>T at a MAF of 24.2% across all studied populations (42.0% in Asians), and resulted in original *CYP2D6*10* allele frequencies in the AA, Asian, Caucasian, Hispanic and AJ of 5.6, 35.6, 3.6, 0.6 and 8.3%, respectively. However, after applying our exon 9 conversion PCR assay to all subjects with a *CYP2D6*10*-containing diplotype, almost half of the **10* alleles were redefined to **36* (3.2%), **36+*10* (31.9%) or **36×N+*10* (10.6%). As such, the revised *CYP2D6*10* allele frequencies in our multiethnic cohort were reduced to 4.4, 13.8, 3.0, 0 and 7.8%, respectively.

MLPA testing confirmed all the *CYP2D6* deletion (**5*) and duplication alleles identified by Luminex genotyping; however, 40 (9.4%) multiethnic samples had copy number gains by MLPA (exons 1, 4 and 6) without a gene duplication signal by Luminex. The absence of a copy number gain at the 3′ downstream region in these cases with concurrent 100C>T was supportive of a *CYP2D6*36+*10* tandem allele as the exon 9 conversion likely interferes with MLPA probe hybridization in the 'downstream' copy. However, it is notable that 2 of 42 (4.8%) isolated *CYP2D6*36* allele carriers did not have a 3′ downstream copy number loss by MLPA, which was attributed to variability in the extent of the downstream gene conversion and its capacity to influence MLPA probe hybridization. This is supported by the location of 3′ downstream MLPA probes and the exon 9 (**36*; 597 bp amplicon) PCR primers, as the MLPA probes hybridize 480–503 bp downstream of exon 9, whereas the reverse PCR primer hybridizes only 434–458 bp downstream of exon 9. As such, integrating our exon 9 conversion PCR assay and MLPA analysis enabled the complementary detection of all complex *CYP2D6*36* alleles, which occur in multiethnic populations (5.0% MAF in our cohort) and most frequently among Asians (21.8% MAF in our cohort).

All other samples with copy number gains detected by both Luminex and MLPA were subjected to duplication allele-specific genotyping using an amplicon that specifically amplified the 'upstream' *CYP2D6* copy or copies and a modified Luminex protocol. This approach eliminated the ambiguity in characterizing duplications [13], which was not previously possible with only Luminex genotyping and/or MLPA testing. Assessing the Luminex median fluorescent intensity ratios of heterozygous alleles among duplication-positive samples can provide support for the haplotype of some 'upstream' copies [20]; however, only direct interrogation of 'upstream' amplicons can characterize these alleles with certainty. Importantly, the majority (61.5%) of duplication alleles in our multiethnic cohort were detected among individuals with compound heterozygous sequence variants that defined two haplotypes with different activity scores. This strategy is similar to the previously reported TaqMan genotyping of 'fragment D' duplication-specific amplicons [12]. The data presented in that study and those presented from this investigation underscore the importance and utility of *CYP2D6* duplication allele-specific genotyping, as the direct characterization of 'upstream' copies in these compound heterozygotes enabled more precise activity score and metabolizer phenotype prediction.

Conclusion

As the frequencies of variant *CYP2D6* alleles continue to be elucidated in diverse populations [31], it is increasingly apparent that comprehensive characterization is needed when studying this important polymorphic gene. Despite several clinically relevant examples of *CYP2D6*-mediated drug response [22,33–36], some *CYP2D6* association studies have been inconsistent, which is likely due, at least in part, to discrepant and/or inadequate *CYP2D6* genotyping [37]. As such, our integrated testing of *CYP2D6* using multiple platforms, including novel assays, enabled the identification of alleles that are found in diverse populations, emphasizing the need to expand common genotyping approaches when interrogating non-Caucasian or admixed populations. Moreover, in the absence of phased long-read sequencing [2] and haplotype-specific [4] approaches, the addition of *CYP2D6*36* and duplication allele-specific assays to multiplexed genotyping improves metabolizer phenotype prediction and may facilitate more consistent outcomes among *CYP2D6* pharmacogenetic association studies.

Summary points

***CYP2D6* genotyping & copy number analysis in a multiethnic cohort**

- Multiplexed targeted genotyping and MLPA analysis of 427 African–American, Asian, Caucasian, Hispanic and Ashkenazi Jewish individuals identified the frequencies of 16 variant *CYP2D6* alleles and the copy number of exons 1, 4, 6 and 3′ downstream region.

Detection of *CYP2D6*36* alleles

- A novel PCR strategy was developed to distinguish the *CYP2D6*36* exon 9 conversion, which was applied to all subjects that carried **10* based on targeted genotyping.
- The combined results of this PCR strategy redefined approximately 46% of all identified **10* alleles into **36* (3.2%), **36*+**10* (31.9%) or **36*×2+**10* (10.6%).
- MLPA analysis across the *CYP2D6* gene supported the results of the *CYP2D6*36* exon 9 conversion PCR assay.

***CYP2D6* duplication allele-specific genotyping**

- Duplication allele-specific multiplexed genotyping unambiguously identified the haplotype of each duplicated *CYP2D6* copy among all tested samples, including samples that were compound heterozygous for *CYP2D6* alleles with different activity scores.

Revised *CYP2D6* allele & genotype frequencies

- The combined allele frequencies after integrating all testing results in the African–American, Asian, Caucasian, Hispanic and Ashkenazi Jewish populations were distributed as increased function (0.038, 0.011, 0.012, 0.024 and 0.104), normal function (0.475, 0.426, 0.596, 0.725 and 0.438), decreased function (0.300, 0.432, 0.175, 0.116 and 0.240) and no function (0.188, 0.120, 0.217, 0.128 and 0.219), respectively.

***CYP2D6* phenotype prediction**

- *CYP2D6* metabolizer phenotypes were inferred based on identified diplotypes and classified using two systems: activity score of 1 as normal metabolizers, and activity score of 1 as intermediate metabolizers.

Conclusion

- Integrated testing of *CYP2D6* using novel genotyping assays and multiple platforms enabled the identification of allele frequencies across diverse populations.
- The addition of *CYP2D6*36* and duplication allele-specific assays to multiplexed genotyping improves metabolizer phenotype prediction and may facilitate more consistent outcomes among *CYP2D6* pharmacogenetic association studies.

Financial & competing interests disclosure

The CYP450 multiplex ligation-dependent probe amplification kit reagents used in this study were generously provided by MRC-Holland (Amsterdam, The Netherlands) and the xTAG CYP2D6 v3 Kits by Luminex Corporation (TX, USA). R Vijzelaar is a paid employee of MRC-Holland, Amsterdam, The Netherlands; W Qiao, G Mendiratta, L Shi, L Edelmann, R Kornreich and SA Scott are paid employees of Sema4, Stamford, CT, USA. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1. Owen RP, Sangkuhl K, Klein TE, Altman RB. Cytochrome P450 2D6. *Pharmacogenet. Genomics* 19(7), 559–562 (2009).
2. Qiao W, Yang Y, Sebra R *et al.* Long-read single molecule real-time full gene sequencing of cytochrome P450-2D6. *Hum. Mutat.* 37(3), 315–323 (2016).
3. Yang Y, Botton MR, Scott ER, Scott SA. Sequencing the *CYP2D6* gene: from variant allele discovery to clinical pharmacogenetic testing. *Pharmacogenomics* 18(7), 673–685 (2017).
4. Gaedigk A, Riffel AK, Leeder JS. CYP2D6 haplotype determination using long range allele-specific amplification: resolution of a complex genotype and a discordant genotype involving the *CYP2D6*59* allele. *J. Mol. Diagn.* 17(6), 740–748 (2015).
5. Nofziger C, Paulmichl M. Accurately genotyping *CYP2D6*: not for the faint of heart. *Pharmacogenomics* 19(13), 999–1002 (2018).
6. Gaedigk A, Ingelman-Sundberg M, Miller NA *et al.* The Pharmacogene Variation (PharmVar) consortium: incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clin. Pharmacol. Ther.* 103(3), 399–401 (2018).
7. Pratt VM, Zehnbauer B, Wilson JA *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.* 12(6), 835–846 (2010).
8. Gaedigk A. Complexities of *CYP2D6* gene analysis and interpretation. *Int. Rev. Psychiatry* 25(5), 534–553 (2013).
9. Pratt VM, Everts RE, Aggarwal P *et al.* Characterization of 137 genomic DNA reference materials for 28 pharmacogenetic genes: a GeT-RM Collaborative Project. *J. Mol. Diagn.* 18(1), 109–123 (2016).
10. 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.* 34(4), 563–569 (2006).
11. Soyama A, Saito Y, Kubo T *et al.* Sequence-based analysis of the *CYP2D6*36-CYP2D6*10* tandem-type arrangement, a major *CYP2D6*10* haplotype in the Japanese population. *Drug Metab. Pharmacokinet.* 21(3), 208–216 (2006).
12. Gaedigk A, Ndjountche L, Divakaran K *et al.* Cytochrome P4502D6 (*CYP2D6*) gene locus heterogeneity: characterization of gene duplication events. *Clin. Pharmacol. Ther.* 81(2), 242–251 (2007).
13. Ramamoorthy A, Skaar TC. Gene copy number variations: it is important to determine which allele is affected. *Pharmacogenomics* 12(3), 299–301 (2011).
14. Gaedigk A, Simon SD, Pearce RE, Bradford LD, Kennedy MJ, Leeder JS. The CYP2D6 activity score: translating genotype information into a qualitative measure of phenotype. *Clin. Pharmacol. Ther.* 83(2), 234–242 (2008).
15. Gaedigk A, Dinh JC, Jeong H, Prasad B, Leeder JS. Ten years' experience with the CYP2D6 activity score: a perspective on future investigations to improve clinical predictions for precision therapeutics. *J. Pers. Med.* 8(2), pii:E15 (2018).
16. Martis S, Peter I, Hulot JS, Kornreich R, Desnick RJ, Scott SA. Multi-ethnic distribution of clinically relevant *CYP2C* genotypes and haplotypes. *Pharmacogenomics J.* 13(4), 369–377 (2013).
17. Scott SA, Edelmann L, Kornreich R, Desnick RJ. Warfarin pharmacogenetics: *CYP2C9* and *VKORC1* genotypes predict different sensitivity and resistance frequencies in the Ashkenazi and Sephardi Jewish populations. *Am. J. Hum. Genet.* 82(2), 495–500 (2008).
18. Scott SA, Edelmann L, Liu L, Luo M, Desnick RJ, Kornreich R. Experience with carrier screening and prenatal diagnosis for 16 Ashkenazi Jewish genetic diseases. *Hum. Mutat.* 31(11), 1240–1250 (2010).
19. Scott SA, Martis S, Peter I, Kasai Y, Kornreich R, Desnick RJ. Identification of *CYP2C19*4B*: pharmacogenetic implications for drug metabolism including clopidogrel responsiveness. *Pharmacogenomics J.* 12(4), 297–305 (2012).
20. Scott SA, Edelmann L, Kornreich R, Erazo M, Desnick RJ. *CYP2C9*, *CYP2C19* and *CYP2D6* allele frequencies in the Ashkenazi Jewish population. *Pharmacogenomics* 8(7), 721–730 (2007).
21. Martis S, Mei H, Vijzelaar R, Edelmann L, Desnick RJ, Scott SA. Multi-ethnic cytochrome-P450 copy number profiling: novel pharmacogenetic alleles and mechanism of copy number variation formation. *Pharmacogenomics J.* 13(6), 558–566 (2013).
22. Goetz MP, Sangkuhl K, Guchelaar HJ *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for *CYP2D6* and tamoxifen therapy. *Clin. Pharmacol. Ther.* 103(5), 770–777 (2018).

23. Bush WS, Crosslin DR, Owusu-Obeng A *et al.* Genetic variation among 82 pharmacogenes: the PGRNseq data from the eMERGE network. *Clin. Pharmacol. Ther.* 100(2), 160–169 (2016).
24. Ng D, Hong CS, Singh LN, Johnston JJ, Mullikin JC, Biesecker LG. Assessing the capability of massively parallel sequencing for opportunistic pharmacogenetic screening. *Genet. Med.* 19(3), 357–361 (2017).
25. Rasmussen-Torvik LJ, Almoguera B, Doheny KF *et al.* Concordance between research sequencing and clinical pharmacogenetic genotyping in the eMERGE-PGx study. *J. Mol. Diagn.* 19(4), 561–566 (2017).
26. Johansson I, Oscarson M, Yue QY, Bertilsson L, Sjoqvist F, Ingelman-Sundberg M. Genetic analysis of the Chinese cytochrome P4502D locus: characterization of variant *CYP2D6* genes present in subjects with diminished capacity for debrisoquine hydroxylation. *Mol. Pharmacol.* 46(3), 452–459 (1994).
27. Yokota H, Tamura S, Furuya H *et al.* Evidence for a new variant *CYP2D6* allele *CYP2D6J* in a Japanese population associated with lower *in vivo* rates of sparteine metabolism. *Pharmacogenetics* 3(5), 256–263 (1993).
28. Gough AC, Miles JS, Spurr NK *et al.* Identification of the primary gene defect at the cytochrome P450 CYP2D locus. *Nature* 347(6295), 773–776 (1990).
29. Sim SC, Ingelman-Sundberg M. Update on allele nomenclature for human cytochromes P450 and the Human Cytochrome P450 Allele (CYP-allele) Nomenclature Database. *Methods Mol. Biol.* 987, 251–259 (2013).
30. Lek M, Karczewski KJ, Minikel EV *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536(7616), 285–291 (2016).
31. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. *Genet. Med.* 19(1), 69–76 (2017).
32. Suwannasri P, Thongnoppakhun W, Pramyothin P, Assawamakin A, Limwongse C. Combination of multiplex PCR and DHPLC-based strategy for *CYP2D6* genotyping scheme in Thais. *Clin. Biochem.* 44(13), 1144–1152 (2011).
33. Bell GC, Caudle KE, Whirl-Carrillo M *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for *CYP2D6* genotype and use of ondansetron and tropisetron. *Clin. Pharmacol. Ther.* 102(2), 213–218 (2017).
34. Hicks JK, Sangkuhl K, Swen JJ *et al.* Clinical pharmacogenetics implementation consortium guideline (CPIC) for *CYP2D6* and *CYP2C19* genotypes and dosing of tricyclic antidepressants: 2016 update. *Clin. Pharmacol. Ther.* 102(1), 37–44 (2017).
35. Hicks JK, Bishop JR, Sangkuhl K *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for *CYP2D6* and *CYP2C19* genotypes and dosing of selective serotonin reuptake inhibitors. *Clin. Pharmacol. Ther.* 98(2), 127–134 (2015).
36. Crews KR, Gaedigk A, Dunnenberger HM *et al.* Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450 2D6 genotype and codeine therapy: 2014 update. *Clin. Pharmacol. Ther.* 95(4), 376–382 (2014).
37. Hertz DL, Mcleod HL, Irvin WJ Jr. Tamoxifen and CYP2D6: a contradiction of data. *Oncologist* 17(5), 620–630 (2012).