Genotype and Phenotype Concordance for Pharmacogenetic Tests Through Proficiency Survey Testing

An Update

Ann M. Moyer, MD, PhD; Gwendolyn A. McMillin, PhD; Thomas A. Long, MPH; Manish J. Gandhi, MD; Rong Mao, MD; Kristi J. Smock, MD; Jaimie G. Halley, MLS(ASCP)CM; Karen E. Weck, MD

• ***Context.***—**As pharmacogenetic testing is incorporated into routine care, it is essential for laboratories to provide accurate and consistent results. Certified laboratories must successfully complete proficiency testing.**

Objectives.—**To understand and examine trends in participation and performance of laboratories participating in the College of American Pathologists pharmacogenetic proficiency testing surveys.**

Design.—**Results from College of American Pathologists pharmacogenetic proficiency testing challenges from 2012 through 2017 were reviewed for concordance with expected genotype and phenotype for each sample (intended responses).**

Results.—**Laboratories correctly reported results for 96.7% to 100% of samples with no variants. Excluding *CYP2D6*, laboratories correctly detected and reported variant alleles for each gene (93.7%–99.2% correct). *CYP2D6* showed lower concordance, with 83.1% of laboratories reporting the intended genotype across all samples; however, in many cases, the laboratories that did not report a variant allele did not test for that allele. Among laboratories reporting the intended genotype, most successfully reported the intended phenotype (85.9%–99.0%).**

Conclusions.—**Although laboratories are generally performing well, there is room for additional improvement, particularly for challenging genes, such as *CYP2D6*. Efforts in the field of pharmacogenomics to recommend alleles that should be included in clinical tests, identify reference materials, and standardize translation from genotype to phenotype may address some of the remaining variability in results.**

(*Arch Pathol Lab Med.* 2020;144:1057–1066; doi: 10.5858/arpa.2019-0478-CP)

The application of genetic data to inform drug and dose selection decisions, known as pharmacogenetics (PGx), has received a growing level of attention from patients, wellness proponents, payers, regulatory bodies, and providers. Pharmacogenetic information is included in the US Food and Drug Administration labeling, as well as international drug labeling, for a subset of prescribed medications.[1] In addition, organizations such as the Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Dutch Pharmacogenomics Working Group have written evidence-based guidelines for the use of PGx test results to facilitate incorporation of this information into clinical practice.[2–6] Some commercial laboratories also use proprietary clinical decision support to apply PGx information to drug and dose selection decisions.[7] Pharmacogenetic information is increasingly incorporated into clinical study design and routine clinical practice with the hope of decreasing adverse medication reactions and selecting the most effective therapy for individuals.

As adoption of PGx increases, efforts to standardize clinical testing and reporting have begun. For example, a working group of the Association for Molecular Pathology (AMP) has published manuscripts recommending a minimum set of alleles that should be included in clinical *CYP2C19* and *CYP2C9* tests, and similar work for other genes is ongoing,[8,9] and the Centers for Disease Control and Prevention Genetic Testing References Materials Coordination Program has identified reference materials that can be used by laboratories during assay validation and as controls when implementing new tests.[10–12] Professional consensus guidelines for nomenclature of variants, assignment of haplotypes and phenotypes, and reporting have also been published.[13–15] Although these efforts to standardize testing and reporting are important, they do not address testing accuracy, which is also of great importance in clinical testing.

Under the Clinical Laboratory Improvement Amendments of 1988 statute, moderate- and high-complexity laboratories are required to verify accuracy twice per year for analytes and

Accepted for publication January 25, 2020.
Published online March 9, 2020.

From the Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota (Moyer, Gandhi); the Department of Pathology and ARUP Laboratories, University of Utah School of Medicine, Salt Lake City (McMillin, Mao, Smock); Biostatistics (Long) and Proficiency Testing (Halley), College of American Pathologists, Northfield, Illinois; and the Department of Pathology and Laboratory Medicine and Department of Genetics, University of North Carolina, Chapel Hill (Weck).

The authors have no relevant financial interest in the products or companies described in this article.

Corresponding author: Ann M. Moyer, MD, PhD, 200 First St SW, Rochester, MN 55905 (email: Moyer.Ann@Mayo.edu).

Table 1. Summary of the Most Common Clinical Applications of Testing by Gene, Along With Examples of Relevant Medications Associated With Each Gene

Gene	Clinical Application of Testing	Commonly Associated Drug(s)
CYP2C19	Predict drug metabolizing efficiency; guide drug and dose selection	Clopidogrel, antidepressants, proton pump inhibitors
CYP2C9	Predict drug metabolizing efficiency; guide drug and dose selection	Warfarin, phenytoin
CYP2D6	Predict drug metabolizing efficiency; guide drug and dose selection	Antidepressants, antipsychotics, atomoxetine, codeine, metoprolol, ondansetron, tamoxifen, tropisetron
*HLA-B*57:01*	Predict likelihood of toxicity; guide drug selection	Abacavir
IL28B (IFNL3)	Predict likelihood of response; guide drug selection	Peginterferon alfa-2, ribavirin
UGT1A1	Predict drug metabolizing efficiency; guide drug and dose selection	Atazanavir, irinotecan
VKORC1	Predict sensitivity to warfarin; guide dose selection	Warfarin

tests for which participation in a proficiency testing (PT) program is not specifically required.[16] Pharmacogenetic tests are not included in the list of Clinical Laboratory Improvement Amendments–regulated analytes for which PT must be performed; however, accrediting agencies may require PT for additional analytes.[17] The College of American Pathologists (CAP), which is an accrediting agency, offers PT surveys for PGx to enable laboratories around the globe to routinely assess analytical performance and ensure accuracy or improve performance when necessary, in order to promote the highest level of patient care. Although many CAP-accredited laboratories that perform PGx testing choose to participate in the CAP PT program, it is not required, and alternative performance assessment methods to ensure accuracy may be used by some laboratories.

The first CAP PGx survey was conducted in 2007 with 2 annual mailings that included 2 samples (referred to as challenges) for 5 genes, as shown in Figure 1. Over time the number of challenges has increased to 3 per mailing, and there are now 4 surveys that cover 12 pharmacogenes. For each gene, the participant reports the genotype and phenotype results for each sample. Additional questions regarding clinical management are included for some genes as well. As of 2017 there were approximately 260 subscribers, including international and US-based laboratories. Although some laboratories participate in all aspects of the surveys, some do not test for all of the genes included. A prior report evaluated the degree of concordance among laboratories during the 2007–2011 time period for 5 genes.[18] Here, we continue this analysis with data from the 2012 through 2017 mailings for 7 of the genes represented in this survey data set: *CYP2C19, CYP2C9, CYP2D6, HLA-B*57:01, IL28B, UGT1A1,* and *VKORC1*. A summary of the most common reasons for clinical testing and examples of relevant medications is shown in Table 1. Proficiency testing performance for genes added to the survey after 2015 is not discussed in detail here because there are insufficient data for analysis. Laboratory performance on questions related to patient management clinical scenarios is also not discussed, because this component of the surveys is included for educational purposes and is currently not graded.

METHODS

In this study, data from CAP PGx PT surveys were evaluated from 2012 through 2017. The CAP PT surveys have been designed to focus on the most commonly tested alleles of genes for which well-characterized reference materials are available. Each laboratory's test results for unknown specimens are compared with the results obtained by other laboratories. This is an ongoing measure of laboratory quality. For each survey challenge, each participating laboratory was provided with a sample containing approximately 25 μg of DNA, selected by the CAP Pharmacogenetics Project Team and purchased from the National Institute of General Medical Sciences Repository at the Coriell Cell Repositories (Camden, New Jersey). Data collected by the CAP from each laboratory participant included methodology (technique and reagent) used by each laboratory for testing, the specific alleles that the laboratory can detect, and the laboratory's results for each challenge (genotype, phenotype prediction, and clinical management related to a clinical scenario, typically corresponding to a CPIC guideline when applicable).

As part of the PT program, the CAP graded responses for intended genotype and phenotype for which there was 80% concordance; those responses that were not graded were considered "educational." For the PT survey, participants received an "acceptable" grade if they failed to detect an allele because it was not included in their assay design. In contrast, for the purposes of this report, we reanalyzed all results and considered responses acceptable only when the intended genotype was reported, regardless of assay design. Genotypes were reported as either star alleles, based on nomenclature historically curated for the cytochrome P450 (*CYP*) genes by the Karolinska Institute[19] and now curated by PharmVar[13]; individual single-nucleotide variants (SNVs); or positive/negative status for a specific human leukocyte antigen (HLA) allele, as appropriate for the gene analyzed. In order to assess the accuracy and consistency of phenotype interpretation among laboratories, phenotype responses were evaluated in this report only for laboratories that provided the intended genotype (ie, if the laboratory reported an incorrect genotype, the reported phenotype would be expected to be incorrect; therefore, this result was excluded).

All summarizations and analyses were completed using SAS (Cary, North Carolina). Chi-square tests compared differences in genotype performance relative to the defined baseline genotype. An omnibus test was also performed comparing overall performance by gene as well as pairwise comparisons of overall performance by gene. Fisher exact tests were used if any cell count was lower than 5. A Bonferroni correction factor was used for statistical significance because of multiple comparisons within each gene; statistical significance was defined as $P < .05$ divided by the number of comparisons for each gene.

RESULTS

An increase in the number of participants for the CAP PGx surveys was noted in the 2014–2015 time frame for the *CYP2C19*, *CYP2C9*, *CYP2D6*, and *VKORC1* challenges (Figure 2). *IL28B* and *HLA-B*57:01* were added to the survey in 2014, and participation has remained fairly stable for *HLA-B*57:01* but has declined for *IL28B* and *UGT1A1*.

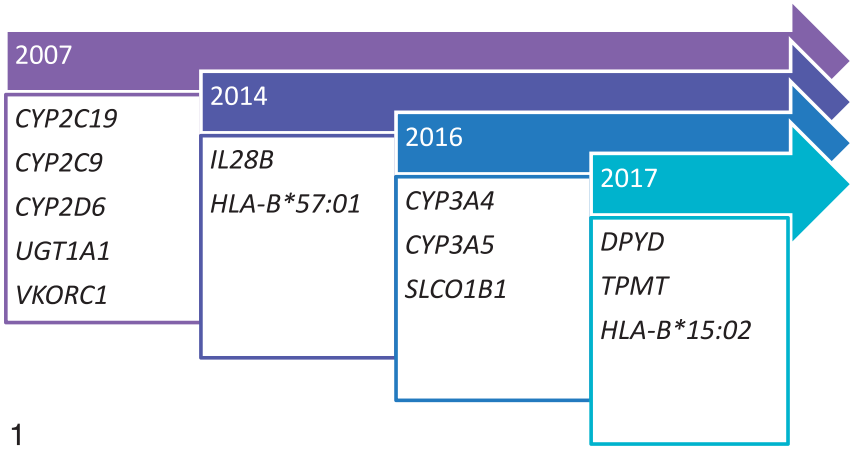

Figure 1. *Genes included in the College of American Pathologists pharmacogenetic proficiency testing surveys from 2007 through 2017.*

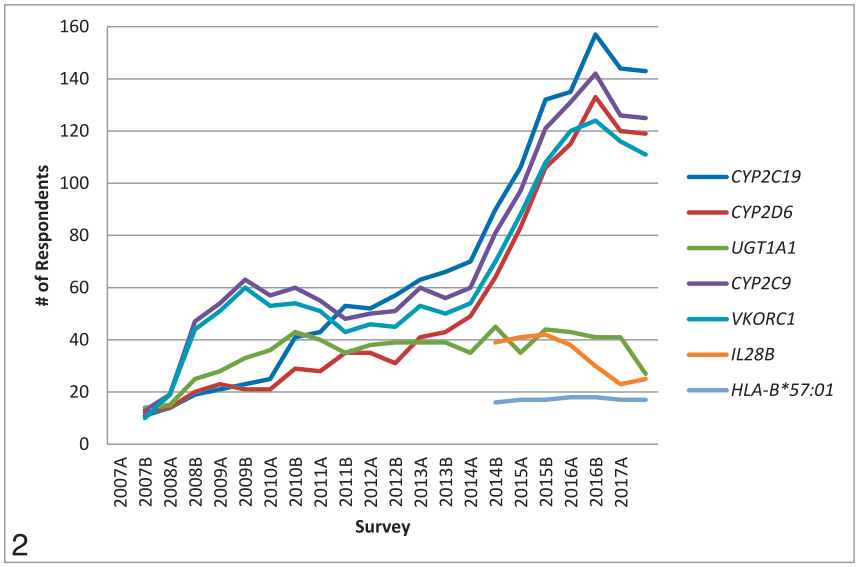

Figure 2. *Average number of participants responding to each pharmacogenetic survey from 2007 through 2017.*

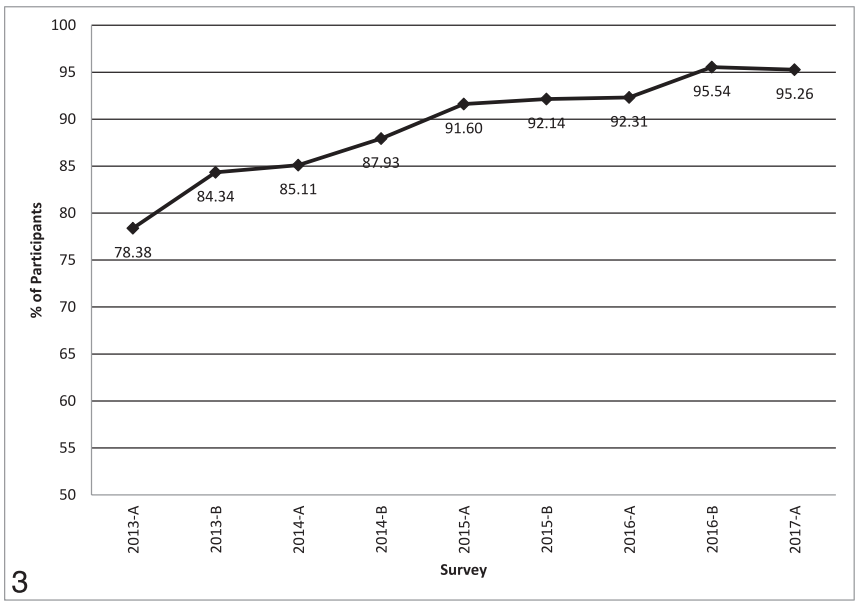

Figure 3. *Percentage of laboratories that reported including the* CYP2C19*17 *allele in their* CYP2C19 *test design over time.*

Laboratory participants reported using a variety of testing technologies including polymerase chain reaction–based genotyping techniques, microarray, Sanger sequencing, and massively parallel (next-generation) sequencing. Notably, the alleles tested and reported for some genes (particularly the *CYP* genes) varied among laboratories.

Results from 2012 through 2017 for each gene are presented in Table 2. In addition to wild-type samples with

Table 2. Genotype and Phenotype Concordance by Gene and Specific Genotype[a]

Gene	Genotype	No. of Challenges	No. Correct/ Total	Genotype Concordance, %	*P* Versus WT[b]	Phenotype	Phenotype Concordance, %	Correct/Reported With Correct Genotype
CYP2C19	**1/*1*	6	739/750	98.5	Baseline	Normal (extensive) metabolizer	95.9	685/714
	**1/*2*	4	630/644	97.8	.32	Intermediate metabolizer	96.9	589/608
	**1/*17*	3	400/433	92.4	*<.001*	Rapid (ultrarapid) metabolizer	94.5	364/385
	**2/*2*	1	154/157	98.1	.72	Poor metabolizer	100.0	146/146
	**2/*3*	2	194/201	96.5	.06	Poor metabolizer	100.0	190/190
	**2/*4*	1	46/74	62.2	*<.001*	Poor metabolizer	97.7	43/44
	**2/*17*	2	288/305	94.4	*.001*	Intermediate metabolizer	94.6	261/276
	**8/*17*	1	50/83	60.2	*<.001*	Intermediate metabolizer	58.0	29/50
	**17/*17*	2	157/190	82.6	*<.001*	Ultrarapid metabolizer	98.0	148/151
	Overall	**22**	**2658/2837**	93.7	...		95.7	**2455/2564**
CYP2C9	**1/*1*	11	1258/1276	98.6	Baseline	Normal (extensive) metabolizer	97.5	1153/1183
	**1/*2*	5	489/503	97.2	.05	Intermediate metabolizer	91.6	423/462
	**1/*3*	2	211/220	95.9	*.006*	Intermediate metabolizer	79.3	157/198
	**1/*5*	1	89/130	68.5	*<.001*	Intermediate metabolizer	95.3	82/86
	**2/*2*	2	212/224	94.6	*<.001*	Poor (slow) metabolizer	87.1	175/201
	**3/*3*	1	69/73	94.5	.03	Poor (slow) metabolizer	100.0	66/66
	Overall	**22**	**2328/2426**	96.0	...		93.6	**2056/2196**
CYP2D6	**1/*1*	2	203/210	96.7	Baseline	Normal (extensive) metabolizer	97.9	184/188
	**1/*2*	2	91/97	93.8	.25	Normal (extensive) metabolizer	97.7	86/88
	**1/*4*	3	298/326	91.4	.02	Normal (extensive) metabolizer	74.9	212/283
	**1/*17*	1	111/122	91.0	.03	Normal (extensive) metabolizer	81.0	81/100
	**1/*29*	1	99/122	81.1	*<.001*	Normal (extensive) metabolizer	81.5	75/92
	**2/*4*	1	60/67	89.6	.02	Normal (extensive) metabolizer	61.4	35/57
	**2/*5*	1	106/144	73.6	*<.001*	Normal (extensive) metabolizer	75.2	76/101
	**2/*10*	2	130/146	89.0	.004	Normal (extensive) metabolizer	85.4	105/123
	**2/*35*	2	117/244	48.0	*<.001*	Normal (extensive) metabolizer	96.4	107/111
	**2/*41*	1	109/116	94.0	.03	Normal (extensive) metabolizer	94.1	96/102
	**4/*4*	1	98/102	96.1	.75	Poor metabolizer	100.0	97/97
	**4/*5*	1	112/143	78.3	*<.001*	Poor metabolizer	99.1	106/107
	**4/*10*	2	126/136	92.6	.09	Intermediate metabolizer	82.0	100/122

Table 2. Continued

Gene	Genotype	No. of Challenges	No. Correct/ Total	Genotype Concordance, %	*P* Versus WT[b]	Phenotype	Phenotype Concordance, %	Correct/Reported With Correct Genotype
	**4/*35*	1	15/54	27.8	*<.001*	Normal (extensive) metabolizer	80.0	12/15
	**1/*2xN* (duplication)	1	112/121	92.6	.09	Ultrarapid metabolizer	76.7	79/103
	Overall	**22**	**1787/2150**	83.1	...		85.9	**1451/1689**
HLA-B	**57:01* absent/ negative	10	182/182	100.0	Baseline	No increased risk of abacavir hypersensitivity reaction (decreased risk of hypersensitivity)	100.0	182/182
	**57:01* present/ positive	11	196/199	98.5	.25	Increased risk of abacavir hypersensitivity reaction	100.0	196/196
	Overall	**21**	**378/381**	99.2	...		100.0	**378/378**
IL28B	rs12979860 C/C	6	216/220	98.2	Baseline	More likely to have SVR	99.5	196/197
	rs12979860 C/T	8	293/296	99.0	.47	Less likely to have SVR	98.1	261/266
	rs12979860 T/T	7	244/248	98.4	.99	Less likely to have SVR	99.5	220/221
	Overall	**21**	**753/764**	98.6	...		99.0	**677/684**
UGT1A1	**1/*1*	5	217/219	99.1	Baseline	Normal metabolizer (standard risk of toxicity)	100.0	208/208
	**1/*28*	10	449/469	95.7	.02	Intermediate metabolizer (increased risk of toxicity)	97.9	412/421
	**28/*28*	8	348/358	97.2	.15	Poor metabolizer (significantly increased risk of toxicity)	99.4	332/334
	Overall	**23**	**1014/1046**	96.9	...		98.9	**952/963**
VKORC1	c.-1639 G/G or 1173 C/C	7	761/790	96.3	Baseline	Low (normal) sensitivity	95.6	695/727
	c.-1639 A/A or 1173 T/T	7	686/702	97.7	.12	High sensitivity	95.6	628/657
	c.-1639 G/A or 1173 C/T	9	800/820	97.6	.15	Medium sensitivity	96.7	742/767
	Overall	**23**	**2247/2312**	97.2	...		96.0	**2065/2151**

Abbreviation: SVR, sustained virologic response.

[a] The number of challenges from 2012–2017 is included along with the percentage of laboratories reporting the intended genotype (genotype concordance) and *P* value for difference from the baseline (**1/*1* or wild type) concordance. The intended phenotype is also provided along with percentage of laboratories reporting the intended phenotype (phenotype concordance), excluding those that provided an incorrect genotype response. In cases where the intended phenotype for a given genotype changed over time, the most current phenotype is provided with previous intended phenotype in parentheses. Overall genotype and phenotype concordance are also provided for each gene.

[b] Italic type indicates a statistically significant difference ($P < .05$) from the baseline (**1/*1* or wild type) concordance.

Table 3. Alternate Phenotypes Reported for Genotypes With Less Than 90% Phenotype Concordance, by Year[a]

Gene	Genotype	Phenotype	Percentage					
			2012	2013	2014	2015	2016	2017
CYP2C19	**8/*17*	Intermediate		58.0				
		Inconclusive		22.0				
		Normal (extensive)		12.0				
		Ultrarapid		8.0				
CYP2C9	**1/*3*	Intermediate			69.9	84.8		
		Poor (slow)			30.1	12.8		
	**2/*2*	Poor (slow)	85.3				87.9	
		Intermediate	14.8				10.7	
CYP2D6	**1/*4*	Normal (extensive)			66.7	77.7	75.8	
		Intermediate			33.3	20.4	24.2	
	**1/*17*	Normal (extensive)						81.0
		Ultrarapid						16.0
	**1/*29*	Normal (extensive)						81.5
		Ultrarapid						15.2
	**2/*4*	Normal (extensive)			61.4			
		Intermediate			36.8			
	**2/*5*	Normal (extensive)					75.3	
		Intermediate					24.8	
	**2/*10*	Normal (extensive)	82.9			86.4		
		Intermediate	17.1			13.6		
	**4/*10*	Intermediate		70.7	87.7			
		Poor		29.3	12.4			
	**4/*35*	Normal (extensive)		80.0				
		Intermediate		20.0				
	**1/*2xN* (duplication)	Ultrarapid					76.7	
		Normal (extensive)					22.3	

[a] Additional phenotypes comprising fewer than 5% of total responses were excluded.

no detectable variants (**1/*1*), the following alleles have been queried in surveys: 5 different *CYP2C19* alleles (**2*, **3*, **4*, **8*, and **17*), 4 *CYP2C9* alleles (**2*, **3*, **5*, and **9*), 9 *CYP2D6* alleles (**2*, **4*, **5* [deletion], **10*, **17*, **29*, **35*, **41*, and gene duplications), presence or absence of the *HLA-B*57:01* haplotype, 1 *IL28B* SNV (rs12979860), the *UGT1A1* promoter TA repeat polymorphism (**28*), and 1 *VKORC1* SNV (NM_014006.5 c.-1639G>A [rs9923231], which by some laboratories is tested by using the tightly linked c.174-136C>T SNV [rs9934438], also known as 1173C>T). For descriptions of the individual variants that comprise each star allele, please refer to the PharmVar Consortium[13] and the HLA Nomenclature[20] Web sites, and for a description of phenotype nomenclature, please see the CPIC publication on standardizing terms for clinical PGx test results.[14]

Overall, laboratories correctly reported the intended genotype response for 93.7% of samples for *CYP2C19*, 96.0% for *CYP2C9*, 83.1% for *CYP2D6*, 99.2% for *HLA-B*57:01*, 98.6% for *IL28B*, 96.9% for *UGT1A1*, and 97.2% for *VKORC1* (Table 2). Statistical significance of variances in allele detection as compared with the wild-type (**1/*1*) genotype for each gene are indicated by *P* values. An omnibus test comparing overall performance per gene ($N > 2000$) revealed statistically significant differences ($P < .001$). Further pairwise testing identified that *CYP2D6* performance differed from *CYP2C19* ($P < .001$), *CYP2C9* ($P < 001$), and *VKORC1* ($P < .001$). One *CYP2D6* sample was not included in the aggregate scoring because of the presence of a rare haplotype that resulted in significantly discordant results depending on the methodology used.[21] One *CYP2C9* specimen was also excluded from the analysis because only 8 of 130 participating laboratories correctly reported the **1/*9* genotype. In this case, the majority of laboratories reported a genotype of *CYP2C9*1/*1* (normal genotype) for this sample, and survey results indicate that only 6% of laboratories offer *CYP2C9*9* testing in their assay panels. However, identification of the *CYP2C9*9* allele has limited clinical impact because this is a normal-function allele, and it is not included in the recently published joint AMP and CAP recommendations for clinical genotyping allele selection[8] as a tier 1 or tier 2 *CYP2C9* allele. Therefore, these data were not included in this analysis.

Among laboratories that reported the correct intended genotype, nearly all also reported the intended phenotype response: overall 95.7% for *CYP2C19*, 93.6% for *CYP2C9*, 85.9% for *CYP2D6*, 100% for *HLA-B*57:01*, 99.0% for *IL28B*, 98.9% for *UGT1A1*, and 96.0% for *VKORC1* (Table 2). Of note, the CAP PGx project team decided to consider the phenotype assignment for *CYP2C19*1/*17* acceptable whether the response was rapid or ultrarapid metabolizer, because of changing recommendations for interpretation of this diplotype as discussed below. A summary of the phenotype responses for those specific genotypes that did not reach 90% consensus is shown in Table 3.

DISCUSSION

We noted an increase in laboratory participation during the 2014–2015 time frame. CPIC began publishing guidelines for the clinical use of PGx test results in 2011 and several years later, large academic centers that were early

Table 4. Potential Reasons for Discordant Results	
Genotype Discordances	**Phenotype Discordances**
Alleles or variants excluded from test design Data entry errors (eg, leaving second allele blank) Sample swaps Detection of a rare genotype Limitations of assay design (eg, additional variants that interfere with primer or probe binding; poor specificity for pseudogene discrimination)	Changes in recommended interpretations Conflicting interpretation in the literature and lack of consensus Uncertainty of enzyme activity

adopters were beginning to publish their studies and results.[22,23] These events may have raised awareness and enthusiasm for PGx, resulting in more laboratories entering this testing space and enrolling in PT. In contrast, participation in *IL28B* genotyping has decreased, potentially because of changes in clinical management of hepatitis C virus, genotype 1. Although it was previously important in therapy, today clinical practice guidelines do not identify pegylated interferon α (PEG-IFNα) as a first-line therapy for most patients with hepatitis C virus, genotype 1.[24] Additionally, participation in *UGT1A1* genotyping has also decreased. This may be in part because *UGT1A1* testing has not been incorporated into current cancer guidelines and is not routine; thus, testing volumes may be too low for some laboratories to maintain this test.[25] For all genes, there is variability among laboratories in terms of the individual SNVs that can be detected because of assay design and the wide variety of methodologies used. However, the majority of laboratories test for the most common alleles and are able to detect them in PT challenges. The majority of the discordances that were observed in this study can be attributed to differences in test design resulting in not all alleles being detected. Other reasons for incorrect responses were entry errors because of leaving the second allele blank on the results form, sample switches with a positive sample from the survey, or reporting of an "other" genotype, which could represent detection of a rare genotype not included in the survey (Table 4). Because of the excellent performance and concordance for both genotype and phenotype for *HLA-B*57:01*, *IL28B*, *UGT1A1*, and *VKORC1*, no further discussion of these genes is provided. Discussions of specific result patterns resulting in less than 80% concordance for *CYP2C19*, *CYP2C9*, and *CYP2D6* are found below.

CYP2C19

There was more than 95% concordance for detection of *CYP2C19*2* and **3*, which are the AMP tier 1 recommended alleles in addition to *CYP2C19*17*.[9] The **2* and **3* alleles are the most common loss-of-function alleles associated with poor metabolism and therefore represent the most clinically relevant alleles for this gene. Over time there was improved performance for detection of *CYP2C19*17* when this allele was included in challenges, with more than 90% of laboratories correctly detecting this allele from 2014 through 2017, compared with 75.7% in the 2013 A mailing (data not shown). For the homozygous genotype *CYP2C19*17/*17*, in the 2013 A mailing, 23.0% of laboratories (17 of 74) called this sample wild-type *CYP2C19*1/*1*; in the 2014 B mailing, only 12.1% (14 of 116) of laboratories called this sample *CYP2C19*1/*1*. In 2016–2017, 6 different challenge samples included the **17* allele (**1/*17* or **2/*17*), which was correctly identified in more than 92% of responses. This represents a continued trend of increased detection of the *CYP2C19*17* allele, with only ∼50% of laboratories detecting **17* (**1/*17*) in the 2010 B mailing and ∼70% in the 2012 A mailing.[18] The increased detection of *CYP2C19*17* correlates with an increasing percentage of laboratories including the **17* allele in their test design during this time period (Figure 3). This is in keeping with recent recommendations that *CYP2C19*17* be included on clinical *CYP2C19* genotyping platforms.[9]

The *CYP2C19*2/*4* and **8/*17* diplotypes were associated with less than 80% genotype concordance (Table 2). Only 60.2% of laboratories correctly reported the *CYP2C19*8/*17* genotype (2013 B mailing), with the majority of laboratories that missed this genotype not detecting the *CYP2C19*8* allele (reported as *CYP2C19*1/*17*) because it was not included in their assay design. Another 2.4% did not report the *CYP2C19*17* allele (reported as **1/*8*), and 13.3% missed both alleles (reported as *CYP2C19*1/*1*). Only 62.2% of laboratories correctly reported the *CYP2C19*2/*4* genotype (2013 A mailing), primarily because they indicated they did not include *CYP2C19*4* on their test platform. Only one laboratory missed the **2* allele and another laboratory mistakenly reported this sample as *CYP2C19*2/*17*. This represents improved performance for detection of *CYP2C19*4* since 2010, when fewer than 50% of laboratories correctly detected this genotype.[18] *CYP2C19*4* is considered a tier 2 allele per the AMP recommendations and is thus optional for clinical laboratories to include in *CYP2C19* genotype testing.[9] Of note, the c.-806C>T (NM_000769.1, rs12248560) promoter SNV that defines the *CYP2C19*17* allele and results in increased activity can be present in cis (on the same allele) with the **4*-defining SNVs, which together define the *CYP2C19*4B* loss-of-function allele; however, determination of the phase of these 2 SNVs is technically challenging and is not possible with most genotyping platforms in clinical use. No *CYP2C19*4B* sample has been included in the CAP survey challenges during this time period.

Although there was more than 95% concordance for phenotypic prediction of most *CYP2C19* genotypes, the *CYP2C19*8/*17* diplotype was associated with only 58.0% concordance for phenotype prediction in the 2013 B mailing (Tables 2 and 3). This is likely due to the historic uncertainty of the enzymatic activity of a null allele in combination with the **17* increased-activity allele and conflicting interpretation in the literature. Additionally, the recommended interpretation of diplotypes that include *CYP2C19*17* has changed during this time frame; it has recently been recommended that **1/*17* be interpreted as rapid metabolizer, **17/*17* as ultrarapid metabolizer, and **17* in combination with a no-function allele, such as **2* or **8*, as intermediate metabolizer.[14] Variability in phenotype responses for *CYP2C19*1/*17* and **8/*17* likely reflects the lack of consensus in the literature and changing recom-

mendations. For example, *CYP2C19*1/*17* reached a concordant (87.6%) interpretation of ultrarapid metabolizer in the 2014 B mailing, weak agreement (56.3%) for rapid metabolizer in the 2016 A mailing, and again greater agreement (70.1%) for rapid metabolizer in the 2017 A mailing (data not shown). This likely reflects recent recommendations that this diplotype be interpreted as a rapid metabolizer instead of an ultrarapid metabolizer.[14,26–28] As mentioned previously, for *CYP2C19*1/*17,* the rapid and ultrarapid phenotype assignments were both considered acceptable for graded CAP PT surveys, so this variability is not apparent in Table 2. The change in majority responses for *CYP2C19*1/*17* phenotype from 2014 to 2017 indicates that laboratories are following changes in recommendations for phenotype interpretation, but that there is some lag in implementing this change universally.

CYP2C9

Performance for detection of the *CYP2C9*2* and **3* reduced-function alleles was 94.5% or better for both heterozygous and homozygous samples. The **2* and **3* alleles are the most common and most clinically relevant *CYP2C9* variant alleles and are currently the only *CYP2C9* alleles included in the US Food and Drug Administration package insert for warfarin dosing. In contrast, in a single survey (2017), only 68.5% of laboratories correctly reported the *CYP2C9*1/*5* diplotype (Table 2). More than 97% of laboratories that did not submit the correct genotype indicated that their assay does not test for the **5* allele, and all erroneous genotypes classified the specimen as **1/*1* (wild type), consistent with inability of the testing platform to detect the **5* allele. Interestingly, approximately 5% of laboratories that reported that their test can detect the **5* allele also submitted an inaccurate genotype, and a single laboratory submitted an accurate genotype despite reporting that their assay does not test for the **5* allele; these discrepancies likely indicate misreporting of alleles covered in the testing platform. Of the laboratories that did submit the correct *CYP2C9*1/*5* genotype, 95% accurately classified the phenotype as an intermediate metabolizer. It should be noted that although some commercial assays detect only the *CYP2C9*2* and **3* alleles, current recommendations are that the **5*, **6*, **8*, and **11* alleles should be included in the assay panel if an African American population is being served because of the frequency and importance of these alleles in warfarin dosing in this population.[29] In addition, these alleles have recently been recommended by AMP and CAP as tier 1 *CYP2C9* alleles that should be included in all *CYP2C9* clinical test platforms.[8] It may take some time for these recommendations to result in changes in clinical practice, given that many laboratories and test platforms currently only include the *CYP2C9*2* and **3* reduced-function alleles.

Although phenotypic interpretation of metabolizer status for *CYP2C9* had overall more than 90% concordance, 2 genotypes had less than 90% phenotype concordance. The *CYP2C9*1/*3* diplotype was associated with 69.7% concordance for phenotype prediction in 2014 and 84.8% concordance in 2015 (Table 3), suggesting improvement in the consensus regarding phenotype assignments for *CYP2C9*. In addition, phenotype concordance for the *CYP2C9*2/*2* diplotype was 85.3% in 2012 and 87.9% in 2016. The *CYP2C9*2* allele is a reduced-function allele with more than 50% residual enzymatic activity, and the stable number of laboratories assigning an intermediate metabolizer instead of poor (slow) metabolizer phenotype for homozygous samples may represent controversy in the field.

CYP2D6

Overall, genotype detection performance for *CYP2D6* was good, with an overall concordance of more than 80%. *CYP2D6* genotyping is challenging because of its highly polymorphic nature as well as the presence of gene deletions (**5*), duplications, and hybrid alleles. Genotype concordance was below 80% for the *CYP2D6*2/*5*, **2/*35*, **4/*5*, and **4/*35* diplotypes (Table 2). The poor concordance when a *CYP2D6*5* or *CYP2D6*35* allele was present was most likely due to laboratories using assays that were not designed to detect these alleles.

In several instances, a laboratory reported an incorrect genotype of *CYP2D6**2/*2 instead of **1/*1*; interestingly, the *CYP2D6* reference sequence in the GRCh37 version of the human genome is a **2A* sequence, which may result in some confusion for laboratories that are not familiar with this gene. Conversely, in several instances, laboratories reported *CYP2D6**2 alleles as **1*. This may again be due to confusion over the reference sequence or due to an assay design that does not include the variant(s) necessary to call a **2* allele. The *CYP2D6*2* allele is thought to have similar to slightly reduced activity as compared with the wild-type **1* allele, so there may be limited clinical significance to miscalling a **2* allele as **1*.

Although the greater than 80% concordance threshold was achieved, it is noteworthy that several laboratories had difficulty distinguishing between *CYP2D6*4* and **10* alleles, **2* and **4* alleles, or **2* and **10* alleles. Several SNVs are common to the *CYP2D6*2*, **4*, and **10* alleles, such as the c.408G>C (p.Val136=) and c.1457G>C (p.Thr486=) synonymous variants (NM_000106.4); however, each of these alleles can be distinguished from one another based on additional variants. For example, the c.100C>T (p.Pro34Ser) variant is present in the *CYP2D6*4* and **10* alleles, but not in **2*. In addition to the c.100C>T SNV, the *CYP2D6*4* allele includes the c.506-1G>A variant, which results in defective splicing and is absent in the *CYP2D6*2* and **10* alleles. Laboratories may have had difficulty in distinguishing these alleles because of assay design that did not include the relevant SNVs, calling haplotypes based on only one variant without consideration of haplotype structure, or other reasons. There were no clear trends in terms of the methodology used for genotyping among the laboratories reporting incorrect results.

The *CYP2D6*35* allele was not detected by some laboratories that incorrectly assigned the **2* genotype for these samples. Review of the data revealed that in each of the 3 challenges that included a sample with the *CYP2D6*35* allele, the majority (>75%) of laboratories that failed to correctly identify the **35* allele reported that they do not test for this allele. Several samples used in early challenges originally had an intended response including a *CYP2D6*2* allele. At that time, the *CYP2D6* c.31G>A, p.Val11Met variant (NM_000106.4, rs769258) associated with *CYP2D6*35* was not included in the design of most assays. Later, after incorporation of this variant into more assays, a subset of the alleles previously identified as *CYP2D6*2* were also found to have the c.31G>A variant and were reclassified as *CYP2D6*35* alleles.

There is also variation in whether laboratories differentiate the *CYP2D6*2A* allele, which includes the c.-1589G>C

promoter variant, from other *2 suballeles. Although some literature suggests that the *CYP2D6* c.-1589G>C variant results in increased transcription and increased activity, other literature disputes this. Therefore, the clinical significance of differentiating between the *CYP2D6*2A* and other *2 alleles is unclear.

CYP2D6 deletion and duplication reporting remains challenging. Approximately 20% of laboratories did not detect the *CYP2D6*5* gene deletion on recent surveys. Again, there were no clear trends in methodology used nor commercial assays associated with failure to detect the *CYP2D6*5* gene deletion. Since 2012, the percentage of laboratories that report that they cannot detect the *CYP2D6*5* gene deletion has increased from 4.6% to 24.6%; however, approximately 10% of the laboratories successfully reported the intended *5 genotype despite indicating that they do not test for this allele. The reason for this discrepancy is unclear.

Reporting a *CYP2D6* gene duplication has also presented challenges. Initially the result form for PT testing included "gene duplication" as a choice along with star alleles; if a duplication was detected, the laboratories were to report duplication rather than the specific star allele. More recently the result form was modified to allow for both allele genotypes to be reported along with an additional separate field to indicate the presence or absence of a duplication. The PT survey instructions and result form have not clearly specified whether a hybrid allele should be reported as a duplication or not. The presence of a hybrid allele may result in discrepancies in reporting of gene duplications if laboratories test for a duplication in only one region of the gene. For example, one sample previously thought to have a *CYP2D6* genotype of *2/*10 was reported by several laboratories to have a gene duplication or "other" allele and later was found to have a *36 *CYP2D6-CYP2D7* hybrid allele. Assays that analyzed the 5′ portion of the gene would have identified a duplication, whereas assays only analyzing the 3′ portion of the gene would have been reported as negative for gene duplication. Assays that detect both regions would have recognized this as a hybrid allele and provided what is now known to be the correct response of *2/*36+*10. Based on the results of supplemental questions included in the 2018 B mailing, the majority of laboratories (61.6% of the 112 laboratories that responded) indicated that they query a single position to test for CNVs, with the majority testing exon 9 in the 3′ portion of the gene. Approximately 53% of laboratories indicated that they would not be able to differentiate between a hybrid allele and a full gene duplication. Accordingly, the aforementioned sample with a genotype of *2/*36+*10 was reported as *2/*10 by some laboratories and *2/gene duplication by other laboratories.

Of the 15 different *CYP2D6* genetic results challenged during this period, phenotype concordance was good (>85%); however, there were 9 diplotypes for which fewer than 90% of participants assigned the intended phenotype (Table 2), which varied by mailing (Table 3). Four of these examples represent the combination of a no-function allele (*4 or *5) and a normal-function allele (*1, *2, or *35). There continues to be controversy over whether this combination introduces clinically relevant impairment of CYP2D6 activity. Historically, this combination of alleles was considered to have CYP2D6 activity falling within the normal range.[30,31] Recent guidelines have recognized that the reduction in CYP2D6 activity in individuals with these diplotypes may be sufficient to have a clinical impact with some CYP2D6 substrates.[32,33] Approximately 25% of laboratory participants classified this combination of alleles as predictive of intermediate metabolizers (Table 3). To help overcome this discrepancy, it is relatively common among laboratories to include an activity score for the predicted CYP2D6 phenotype.[34,35] Activity scores are assigned based on the combination of alleles that are detected, with a no-function allele having an activity score of 0 whereas a normal-function allele has a score of 1. Although an activity score predicted to be 1 (of 2 based on 2 normal-function alleles) might result in sufficient enzymatic activity to have limited impact for some substrates, it may have significance with other substrates and may increase susceptibility to drug-drug interactions or other nongenetic factors that influence dosing.

Two other examples of CYP2D6 phenotype discrepancy in the surveys included a *10 allele, which historically has been considered to be a reduced-activity allele; however, emerging evidence suggests that this allele may have an activity between that of most reduced-activity alleles and null alleles.[33] Therefore, the lack of concordance here may also point to controversy in the field. Another error resulting in less than 90% concordance was observed in the 2017 survey for *CYP2D6*, wherein approximately 15% of laboratories assigned the ultrarapid metabolizer phenotype to samples that included 1 normal-function allele (*1) and 1 decreased-function allele (*17, *29). It is unclear why this misinterpretation occurred. The ninth example involved a *1/*2 sample with gene duplication (*CYP2D6*1/*2xN*). It is unclear why 22.3% of laboratories reported this genotype as a normal (extensive) metabolizer.

Efforts Underway to Improve Standardization

The recent publications lead by AMP that recommend a minimum set of alleles to include in clinical genotyping tests may increase standardization of testing by encouraging laboratories to incorporate the most important alleles into their tests, if they have not already done so.[8,9] Some alleles that are rare in populations of European descent but common among other populations are often excluded from clinical testing. For example, *CYP2C9*5, *6, *8,* and *11 are more common among individuals of African or African American descent, and, if these alleles are not included in clinical testing, individuals with these alleles may be miscalled as normal metabolizers and receive inappropriate dosing recommendations for warfarin. Thus, if laboratories comply with the recommendations for a minimum set of alleles to be included in clinical testing, test results may be more consistent among laboratories, and the potential for introduction of health care disparities by inappropriate test design will be reduced. Although the CAP PGx surveys currently do not penalize laboratories that do not detect an allele not included in their test design, the PGx project team responsible for grading the PT is hopeful that laboratories will follow the recommendations. Similar work involving additional genes beyond *CYP2C19* and *CYP2C9* will be greatly beneficial to the laboratory community.

The majority of laboratories provide similar phenotype interpretations when the same genotypes are identified; however, controversy remains in translation of diplotypes involving certain combinations of allele categories for *CYP2D6*. Work is ongoing that may lead to increased standardization of *CYP2D6* genotype to phenotype translations.[36] The CPIC guidelines for other genes often include

straightforward recommendations for genotype-to-phenotype translation, which may be reflected in the overall high degree of concordance identified in our study.

CONCLUSIONS

In general, we believe that our analysis of the CAP PGx PT data suggests that laboratories are performing well, both in detecting the alleles that they indicate their tests are designed to detect and in assigning phenotypes based on genotyping results. Several alleles that present particular challenges to laboratories have been highlighted, most of which are not detected because of a lack of inclusion in test design.

Trends observed in PT survey results toward increased concordance may demonstrate that the efforts underway by a variety of groups, including AMP, CPIC, and PharmVar, to increase standardization of PGx testing are impacting the laboratory community. Continuation of these efforts along with participation in PT may lead to further harmonization of PGx result reporting.

References

1. US Food and Drug Administration. Table of pharmacogenomic biomarkers in drug labeling. https://www.fda.gov/drugs/science-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling. Accessed June 1, 2019.
2. Whirl-Carrillo M, McDonagh EM, Hebert JM, et al. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther*. 2012;92(4):414–417.
3. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther*. 2011;89(3):464–467.
4. Caudle KE, Klein TE, Hoffman JM, et al. Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process. *Curr Drug Metab*. 2014;15(2):209–217.
5. Swen JJ, Wilting I, de Goede AL, et al. Pharmacogenetics: from bench to byte. *Clin Pharmacol Ther*. 2008;83(5):781–787.
6. Swen JJ, Nijenhuis M, de Boer A, et al. Pharmacogenetics: from bench to byte—an update of guidelines. *Clin Pharmacol Ther*. 2011;89(5):662–673.
7. Bousman CA, Hopwood M. Commercial pharmacogenetic-based decision-support tools in psychiatry. *Lancet Psychiatry*. 2016;3(6):585–590.
8. Pratt VM, Cavallari LH, Del Tredici AL, et al. Recommendations for clinical CYP2C9 genotyping allele selection: a joint recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn*. 2019;21(5):746–755.
9. Pratt VM, Del Tredici AL, Hachad H, et al. Recommendations for clinical CYP2C19 genotyping allele selection: a report of the Association for Molecular Pathology. *J Mol Diagn*. 2018;20(3):269–276.
10. Bettinotti MP, Ferriola D, Duke JL, et al. Characterization of 108 genomic DNA reference materials for 11 human leukocyte antigen loci: a GeT-RM collaborative project. *J Mol Diagn*. 2018;20(5):703–715.
11. Pratt VM, Everts RE, Aggarwal P, et al. Characterization of 137 genomic DNA genomic reference materials for 28 pharmacogenetic genes: a GeT-RM collaborative project. *J Mol Diagn*. 2016;18(1):109–123.
12. 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*. 2010;12(6):835–846.
13. Pharmacogene Variation Consortium (PharmVar). www.PharmVar.org. Accessed June 1, 2019.
14. Caudle KE, Dunnenberger HM, Freimuth RR, et al. Standardizing terms for clinical pharmacogenetic test results: consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC). *Genet Med*. 2017;19(2):215–223.
15. Kalman LV, Agundez J, Appell ML, et al. Pharmacogenetic allele nomenclature: international workgroup recommendations for test result reporting. *Clin Pharmacol Ther*. 2016;99(2):172–185.
16. Participation in proficiency testing for laboratories performing nonwaived testing. 42 CFR §§493.821–493.865. https://www.ecfr.gov/cgi-bin/text-idx?SID=5dc44edd05f61dbc882168bf0af34601&mc=true&node=sp42.5.493.h&rgn=div6. Accessed November 15, 2019.
17. College of American Pathologists, Commission on Laboratory Accreditation. *All Common Checklist*. August 22, 2018 ed. Northfield, IL: College of American Pathologists; 2018.
18. Wu AH. Genotype and phenotype concordance for pharmacogenetic tests through proficiency survey testing. *Arch Pathol Lab Med*. 2013;137(9):1232–1236.
19. Sim SC, Ingelman-Sundberg M. The human cytochrome P450 (CYP) allele nomenclature website: a peer-reviewed database of CYP variants and their associated effects. *Hum Genomics*. 2010;4(4):278–281.
20. Nomenclature for factors of the HLA system. http://hla.alleles.org/nomenclature/index.html. Accessed November 15, 2019.
21. Peterson S, Wimmer A, O'Kane D. A comprehensive full gene sequencing assay for the CYP2D6 gene [abstract G05]. *J Mol Diagn*. 2010;12(6):858.
22. Van Driest SL, Shi Y, Bowton EA, et al. Clinically actionable genotypes among 10 000 patients with preemptive pharmacogenomic testing. *Clin Pharmacol Ther*. 2014;95(4):423–431.
23. Bielinski SJ, Olson JE, Pathak J, et al. Preemptive genotyping for personalized medicine: design of the right drug, right dose, right time-using genomic data to individualize treatment protocol. *Mayo Clin Proc*. 2014;89(1):25–33.
24. AASLD-IDSA HCV Guidance Panel. Hepatitis C guidance 2018 update: AASLD-IDSA recommendations for testing, managing, and treating hepatitis C virus infection. *Clin Infect Dis*. 2018;67(10):1477–1492.
25. de Man FM, Goey AKL, van Schaik RHN, Mathijssen RHJ, Bins S. Individualization of irinotecan treatment: a review of pharmacokinetics, pharmacodynamics, and pharmacogenetics. *Clin Pharmacokinet*. 2018;57(10):1229–1254.
26. PharmGKB. Gene-specific information tables for CYP2C19. https://www.pharmgkb.org/page/cyp2c19RefMaterials. Accessed June 1, 2019.
27. Scott SA, Sangkuhl K, Stein CM, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin Pharmacol Ther*. 2013;94(3):317–323.
28. Moriyama B, Obeng AO, Barbarino J, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for CYP2C19 and voriconazole therapy. *Clin Pharmacol Ther*. 2017;102(1):45–51.
29. Johnson JA, Caudle KE, Gong L, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin Pharmacol Ther*. 2017;102(3):397–404.
30. Crews KR, Gaedigk A, Dunnenberger HM, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype. *Clin Pharmacol Ther*. 2012;91(2):321–326.
31. 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*. 2015;98(2):127–134.
32. Brown JT, Bishop JR, Sangkuhl K, et al. Clinical Pharmacogenetics Implementation Consortium guideline for cytochrome P450 (CYP)2D6 genotype and atomoxetine therapy. *Clin Pharmacol Ther*. 2019;106(1):94–102.
33. Goetz MP, Sangkuhl K, Guchelaar HJ, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for CYP2D6 and tamoxifen therapy. *Clin Pharmacol Ther*. 2018;103(5):770–777.
34. 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*. 2018;8(2). doi:10.3390/jpm8020015
35. Hicks JK, Swen JJ, Gaedigk A. Challenges in CYP2D6 phenotype assignment from genotype data: a critical assessment and call for standardization. *Curr Drug Metab*. 2014;15(2):218–232.
36. Caudle KE, Sangkuhl K, Whirl-Carrillo M, et al. Standardizing CYP2D6 genotype to phenotype translation: consensus recommendations from the Clinical Pharmacogenetics Implementation Consortium and Dutch Pharmacogenomics Working Group [published online October 24, 2019]. *Clin Transl Sci*. 2020;13(1):116–124.