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.

tests for which participation in a proficiency testing (PT) program is not specifically required. Pharmacogenetic tests are not included in the list of Clinical Laboratory Improvement Amendment–regulated analytes for which PT must be performed; however, accrediting agencies may require PT for additional analytes. 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. 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 and now curated by PharmVar; 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.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Clinical Application of Testing</th>
<th>Commonly Associated Drug(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2C19</td>
<td>Predict drug metabolizing efficiency; guide drug and dose selection</td>
<td>Clopidogrel, antidepressants, proton pump inhibitors</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>Predict drug metabolizing efficiency; guide drug and dose selection</td>
<td>Warfarin, phenytoin</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>Predict drug metabolizing efficiency; guide drug and dose selection</td>
<td>Antidepressants, antipsychotics, atomoxetine, codeine, metoprolol, ondansetron, tamoxifen, tropisetron</td>
</tr>
<tr>
<td>HLA-B*57:01</td>
<td>Predict likelihood of toxicity; guide drug selection</td>
<td>Abacavir</td>
</tr>
<tr>
<td>IL28B (IFNL3)</td>
<td>Predict likelihood of response; guide drug selection</td>
<td>Peginterferon alfa-2, ribavirin</td>
</tr>
<tr>
<td>UGT1A1</td>
<td>Predict drug metabolizing efficiency; guide drug and dose selection</td>
<td>Atazanavir, irinotecan</td>
</tr>
<tr>
<td>VKORC1</td>
<td>Predict sensitivity to warfarin; guide dose selection</td>
<td>Warfarin</td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Gene</th>
<th>Genotype</th>
<th>No. of Challenges</th>
<th>No. Correct/Total</th>
<th>Genotype Concordance, %</th>
<th>P Versus WTb</th>
<th>Phenotype</th>
<th>Phenotype Concordance, %</th>
<th>Correct/Reported With Correct Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2C19</td>
<td>*1/*1</td>
<td>6</td>
<td>739/750</td>
<td>98.5</td>
<td>Baseline</td>
<td>Normal (extensive) metabolizer</td>
<td>95.9</td>
<td>685/714</td>
</tr>
<tr>
<td></td>
<td>*1/*2</td>
<td>4</td>
<td>630/644</td>
<td>97.8</td>
<td>.32</td>
<td>Intermediate metabolizer</td>
<td>96.9</td>
<td>589/608</td>
</tr>
<tr>
<td></td>
<td>*1/*17</td>
<td>3</td>
<td>400/433</td>
<td>92.4</td>
<td>&lt;.001</td>
<td>Rapid (ultrarapid) metabolizer</td>
<td>94.5</td>
<td>364/385</td>
</tr>
<tr>
<td></td>
<td>*2/*2</td>
<td>1</td>
<td>154/157</td>
<td>98.1</td>
<td>.72</td>
<td>Poor metabolizer</td>
<td>100.0</td>
<td>146/146</td>
</tr>
<tr>
<td></td>
<td>*2/*3</td>
<td>2</td>
<td>194/201</td>
<td>96.5</td>
<td>.06</td>
<td>Poor metabolizer</td>
<td>100.0</td>
<td>190/190</td>
</tr>
<tr>
<td></td>
<td>*2/*4</td>
<td>1</td>
<td>47/74</td>
<td>62.2</td>
<td>&lt;.001</td>
<td>Intermediate metabolizer</td>
<td>94.6</td>
<td>261/276</td>
</tr>
<tr>
<td></td>
<td>*2/*17</td>
<td>2</td>
<td>288/305</td>
<td>94.4</td>
<td>&lt;.001</td>
<td>Intermediate metabolizer</td>
<td>58.0</td>
<td>29/50</td>
</tr>
<tr>
<td></td>
<td>*17/*17</td>
<td>2</td>
<td>157/190</td>
<td>82.6</td>
<td>&lt;.001</td>
<td>Ultrarapid metabolizer</td>
<td>98.0</td>
<td>148/151</td>
</tr>
<tr>
<td>Overall</td>
<td><strong>22</strong></td>
<td><strong>2658/2837</strong></td>
<td></td>
<td><strong>93.7</strong></td>
<td></td>
<td></td>
<td><strong>95.7</strong></td>
<td><strong>2455/2564</strong></td>
</tr>
<tr>
<td>CYP2C9</td>
<td>*1/*1</td>
<td>11</td>
<td>1258/1276</td>
<td>98.6</td>
<td>Baseline</td>
<td>Normal (extensive) metabolizer</td>
<td>97.5</td>
<td>1153/1183</td>
</tr>
<tr>
<td></td>
<td>*1/*2</td>
<td>5</td>
<td>489/503</td>
<td>97.2</td>
<td>.05</td>
<td>Intermediate metabolizer</td>
<td>91.6</td>
<td>423/462</td>
</tr>
<tr>
<td></td>
<td>*1/*3</td>
<td>2</td>
<td>211/220</td>
<td>95.9</td>
<td>.006</td>
<td>Intermediate metabolizer</td>
<td>79.3</td>
<td>157/198</td>
</tr>
<tr>
<td></td>
<td>*1/*5</td>
<td>1</td>
<td>89/130</td>
<td>68.5</td>
<td>&lt;.001</td>
<td>Intermediate metabolizer</td>
<td>95.3</td>
<td>82/86</td>
</tr>
<tr>
<td></td>
<td>*2/*2</td>
<td>2</td>
<td>212/224</td>
<td>94.6</td>
<td>&lt;.001</td>
<td>Poor (slow) metabolizer</td>
<td>87.1</td>
<td>175/201</td>
</tr>
<tr>
<td></td>
<td>*3/*3</td>
<td>1</td>
<td>69/73</td>
<td>94.5</td>
<td>.03</td>
<td>Poor (slow) metabolizer</td>
<td>100.0</td>
<td>66/66</td>
</tr>
<tr>
<td>Overall</td>
<td><strong>22</strong></td>
<td><strong>2382/2426</strong></td>
<td></td>
<td><strong>96.0</strong></td>
<td></td>
<td></td>
<td><strong>93.6</strong></td>
<td><strong>2056/2196</strong></td>
</tr>
<tr>
<td>CYP2D6</td>
<td>*1/*1</td>
<td>2</td>
<td>203/210</td>
<td>96.7</td>
<td>Baseline</td>
<td>Normal (extensive) metabolizer</td>
<td>97.9</td>
<td>184/188</td>
</tr>
<tr>
<td></td>
<td>*1/*2</td>
<td>2</td>
<td>91/97</td>
<td>93.8</td>
<td>.25</td>
<td>Normal (extensive) metabolizer</td>
<td>97.7</td>
<td>86/88</td>
</tr>
<tr>
<td></td>
<td>*1/*4</td>
<td>3</td>
<td>298/326</td>
<td>91.4</td>
<td>.02</td>
<td>Normal (extensive) metabolizer</td>
<td>74.9</td>
<td>212/283</td>
</tr>
<tr>
<td></td>
<td>*1/*17</td>
<td>1</td>
<td>111/122</td>
<td>91.0</td>
<td>.03</td>
<td>Normal (extensive) metabolizer</td>
<td>81.0</td>
<td>81/100</td>
</tr>
<tr>
<td></td>
<td>*1/*29</td>
<td>1</td>
<td>99/122</td>
<td>81.1</td>
<td>&lt;.001</td>
<td>Normal (extensive) metabolizer</td>
<td>81.5</td>
<td>75/92</td>
</tr>
<tr>
<td></td>
<td>*2/*4</td>
<td>1</td>
<td>60/67</td>
<td>89.6</td>
<td>.02</td>
<td>Normal (extensive) metabolizer</td>
<td>61.4</td>
<td>35/57</td>
</tr>
<tr>
<td></td>
<td>*2/*5</td>
<td>1</td>
<td>106/144</td>
<td>73.6</td>
<td>&lt;.001</td>
<td>Normal (extensive) metabolizer</td>
<td>75.2</td>
<td>76/101</td>
</tr>
<tr>
<td></td>
<td>*2/*10</td>
<td>2</td>
<td>130/146</td>
<td>89.0</td>
<td>.004</td>
<td>Normal (extensive) metabolizer</td>
<td>85.4</td>
<td>105/123</td>
</tr>
<tr>
<td></td>
<td>*2/*35</td>
<td>2</td>
<td>117/244</td>
<td>48.0</td>
<td>&lt;.001</td>
<td>Normal (extensive) metabolizer</td>
<td>96.4</td>
<td>107/111</td>
</tr>
<tr>
<td></td>
<td>*2/*41</td>
<td>1</td>
<td>109/116</td>
<td>94.0</td>
<td>.03</td>
<td>Normal (extensive) metabolizer</td>
<td>94.1</td>
<td>96/102</td>
</tr>
<tr>
<td></td>
<td>*4/*4</td>
<td>1</td>
<td>98/102</td>
<td>96.1</td>
<td>.75</td>
<td>Poor metabolizer</td>
<td>100.0</td>
<td>97/97</td>
</tr>
<tr>
<td></td>
<td>*4/*5</td>
<td>1</td>
<td>112/143</td>
<td>78.3</td>
<td>&lt;.001</td>
<td>Poor metabolizer</td>
<td>99.1</td>
<td>106/107</td>
</tr>
<tr>
<td></td>
<td>*4/*10</td>
<td>2</td>
<td>126/136</td>
<td>92.6</td>
<td>.09</td>
<td>Intermediate metabolizer</td>
<td>82.0</td>
<td>100/122</td>
</tr>
</tbody>
</table>
Table 2. Continued

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

Abbreviation: SVR, sustained virologic response.

<sup>a</sup> 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.

<sup>b</sup> Italic type indicates a statistically significant difference (P < .05) from the baseline (*1/*1 or wild type) concordance.
some laboratories is tested by using the tightly linked c.174-
SNV (NM_014006.5 c.-1639G
CYP2D6
alleles (star allele, please refer to the PharmVar Consortium13 and
descriptions of the individual variants that comprise each
*28
promoter TA repeat polymorphism (B*57:01 haplotype, 1
HLA-
and gene duplications), presence or absence of the
discordant results depending on the methodology used.21
presence of a rare haplotype that resulted in significantly
not included in the aggregate scoring because of the
allele detection as compared with the wild-type (*1/*1
(Table 2). Statistical significance of variances in
genotype for each gene are indicated by
P
001), and
,CYP2D6
Further pairwise testing identified that
2000) revealed statistically significant differences (P
CYP2D6
*1/*4
Normal (extensive) 80.0
*4/*35
Intermediate 20.0
*4/*10
Intermediate 29.3
*4/*35
Normal (extensive) 80.0
*2/*5
Intermediate 36.8
*2/*10
Intermediate 82.9
Intermediate 17.1
*4/*10
Intermediate 70.7
*4/*35
Normal (extensive) 80.0
*2/*4
Normal (extensive) 61.4
*2/*5
Normal (extensive) 75.3
*2/*10
Intermediate 17.1
*4/*10
Intermediate 20.0
*1/*2xN (duplication) Ultrarapid 76.7
Normal (extensive) 22.3
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
*1/*20 Normal (extensive) 86.4
Intermediate 24.8
*4/*11 Intermediate 70.7
*4/*35
Normal (extensive) 87.7
Poor 29.3
*4/*35
Intermediate 13.6
*1/*2xN (duplication) Ultrarapid 16.0
Normal (extensive) 22.3

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

* 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 Consortium13 and
the HLA Nomenclature20 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
phenotype 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.23
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 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
phenotype, 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
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>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2C19</td>
<td>*8/*17</td>
<td>Intermediate</td>
<td>58.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2C9</td>
<td>*1/*3</td>
<td>Intermediate</td>
<td>69.9</td>
<td>84.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2D6</td>
<td>*1/*4</td>
<td>Normal (extensive)</td>
<td>66.7</td>
<td>77.7</td>
<td>75.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*1/*17</td>
<td>Normal (extensive)</td>
<td>81.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*1/*29</td>
<td>Normal (extensive)</td>
<td>81.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*1/*20</td>
<td>Intermediate</td>
<td>24.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*4/*11</td>
<td>Intermediate</td>
<td>70.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*4/*35</td>
<td>Intermediate</td>
<td>20.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*1/*2xN (duplication)</td>
<td>Ultrarapid</td>
<td>76.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal (extensive)</td>
<td>22.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
alleles in addition to
CYP2C19*17
CYP2C19*2
CYP2C19
result patterns resulting in less than 80% concordance for
could represent detection of a rare genotype not included in
represents a continued trend of increased detection of the
correctly identified in more than 92% of responses. This
has not been incorporated into current cancer guidelines
decreased. This may be in part because UGT1A1
representation in the literature and lack of consensus
Uncertainty of enzyme activity

Table 4. Potential Reasons for Discordant Results

| Alleles or variants excluded from test design | Changes in recommended interpretations |
| Data entry errors (eg, leaving second allele blank) | Conflicting interpretation in the literature |
| 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) | |

adaptors were beginning to publish their studies and results. 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. 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. 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 discords 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. 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 (*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. 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.

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. 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. 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 diploype 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. 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 metabo-

lizer 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

Proiciency Testing and Pharmacogenetic Tests—Moyer et al
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 diplostates 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 *3). 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. Approximtely 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. 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. 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. 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 diplostates involving certain combinations of allele categories for CYP2D6. Work is ongoing that may lead to increased standardization of CYP2D6 genotype to phenotype translations. 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