ARTICLE

CYP2C9 and CYP2C19: Deep Mutational Scanning and Functional Characterization of Genomic Missense Variants

Lingxin Zhang1, Vivekananda Sarangi2, Irene Moon1, Jia Yu1, Duan Liu1, Sandhya Devarajan1, Joel M. Reid1, Krishna R. Kalari2, Liewei Wang1,* and Richard Weinshilboum1,*

Single nucleotide variants in the open reading frames (ORFs) of pharmacogenes are important causes of interindividual variability in drug response. The functional characterization of variants of unknown significance within ORFs remains a major challenge for pharmacogenomics. Deep mutational scanning (DMS) is a high-throughput technique that makes it possible to analyze the functional effect of hundreds of variants in a parallel and scalable fashion. We adapted a “landing pad” DMS system to study the function of missense variants in the ORFs of cytochrome P450 family 2 subfamily C member 9 (CYP2C9) and cytochrome P450 family 2 subfamily C member 19 (CYP2C19). We studied 230 observed missense variants in the CYP2C9 and CYP2C19 ORFs and found that 19 of 109 CYP2C9 and 36 of 121 CYP2C19 variants displayed less than ~25% of the wild-type protein expression, a level that may have clinical relevance. Our results support DMS as an efficient method for the identification of damaging ORF variants that might have potential clinical pharmacogenomic application.

Study Highlights

**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**

✔ With the increasing application of next generation sequencing (NGS) to known “pharmacogenes,” large numbers of open reading frame (ORF) “variants of unknown significance” (VUS) are being identified. However, the functional implications of those VUS remains unclear.

**WHAT QUESTION DID THIS STUDY ADDRESS?**

✔ This study was designed to determine whether the application of DMS might make it possible to determine the functional effects of VUS that have been observed in the ORFs of cytochrome P450 family 2 subfamily C member 9 (CYP2C9) and cytochrome P4540 family 2 subfamily C member 19 (CYP2C19).

**WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?**

✔ The results of the study demonstrate that deep mutational scanning (DMS) can be used to determine the functional implications of ORF VUS in CYP2C9 and CYP2C19.

**HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?**

✔ The application of DMS to additional pharmacogenes would potentially expand the accuracy and clinical utility of the application of NGS to pharmacogenomics.

Genetic polymorphisms in or near pharmacogenes are a major cause of individual variation in drug response phenotypes.1 Cytochrome P450 family 2 subfamily C member 9 (CYP2C9) and cytochrome P450 family 2 subfamily C member 19 (CYP2C19) are genes that encode important cytochrome P450 enzymes that catalyze the phase I biotransformation of a variety of therapeutic drugs, including antiplatelet agents, selective serotonin reuptake inhibitors, and proton pump inhibitors.2–5 Several years ago, the Mayo Clinic launched the RIGHT 1K study, in which next generation DNA sequencing (NGS) was performed with DNA from 1013 Mayo Biobank participants to identify variants in 84 pharmacogenes, including CYP2C9 and CYP2C19.1,6 However, many of the polymorphisms observed in those patients were variants of unknown significance (VUS).7–8 In a recent publication, we functionally characterized six novel nonsynonymous open reading frame (ORF) variants in the CYP2C9 gene and seven nonsynonymous ORF variants in the CYP2C19 gene observed in DNA from participants in the Right 1K study.9 Conventional methods for the characterization of individual sequence variants “one-at-a-time” are reliable, but they are also time-consuming, labor-intense, and not easily scalable. As a result, only a limited number of variants can practically be investigated in that fashion. To help address this challenge, predictive algorithms, such as Polyphen-2, SIFT, and PROVEAN, among others, represent efforts to help identify deleterious variants, but their reliability is variable and inadequate for clinical application.10–12

1Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, Minnesota, USA; 2Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. *Correspondence: Richard Weinshilboum (weinshilboum.richard@mayo.edu) and Liewei Wang (Wang.Liewei@mayo.edu).

Received: October 24, 2019; accepted: November 10, 2019. doi:10.1111/cts.12758
Laboratory-based assays are still required to reliably interpret the impact of ORF missense variants on protein function. As a result, a significant gap remains in the functional interpretation of the large number of missense VUS being discovered as DNA sequencing is applied ever more broadly. Deep mutational scanning (DMS) methods provide a platform with which a large number of missense variants can be interrogated in parallel. For example, Fowler et al. developed a DMS assay for all possible variants in clinically important genes, such as BRCA1, TPMT, and PTEN, linking genotypes to functionally determined phenotypes. Specifically, engineered “landing pad” HEK 293T cells were used as a platform to integrate pooled variant libraries, resulting in one variant per cell for further functional assessment. DMS includes the creation of variant libraries for a gene, selecting the library for function (e.g., resistance to drug or fluorescence markers of protein quantity) and—finally—high-throughput DNA sequencing of variants to link them with the “activity” assayed in a functional test.

In our previous “one-at-a-time” functional genomic study, we identified a series of missense variants in the ORFs of CYP2C9 and CYP2C19 that resulted in decreased protein expression as a result of proteasome-mediated degradation, presumably due to an alteration in protein folding. In the present study, fluorescence-activated cell sorting (FACS) was used as a selection mechanism for cells expressing variant sequences (i.e., the cells were sorted according to the abundance of reporter protein expression). We should point out that saturation mutagenesis can also be used to create a “saturation” library, including all possible mutations for an ORF sequence in a single reaction, and that approach has been applied in some previous DMS studies. However, many of the variants created are unlikely to occur clinically. Therefore, we applied a different mutagenesis approach, nicking mutagenesis, to generate libraries that contained human genomic variation known to be present in the general population. Generation of variants by nicking mutagenesis is not exhaustive, but it is much more efficient than the generation of mutants one at a time by site-directed mutagenesis.

In the present study, we have used DMS to analyze the functional implications of missense variants that have been observed in the ORFs of CYP2C9 and CYP2C19, ORFs that were fused to green fluorescent protein (GFP). Recombinase was used to integrate the variant libraries into landing pad cells, one per cell. Multiplexed functional selection performed with FACS was then used to separate the cells into different “bins” based on fluorescence readout at the single cell level. Amplicon sequencing of DNA in each bin, followed by computational analysis of the frequency of variants appearing in each bin, was used to determine levels of protein expression. To identify potentially severely damaging variants for these two important CYP pharmacogenes, we analyzed 230 observed missense variants in ORFs present in a publicly available database consisting of exome sequencing data for 60,000 general population subjects. We included genetic variants with allele frequencies (> 0.00001) observed by the Exome Aggregation Consortium (ExAC; http://exac.broad institute.org/) as well as novel ORF VUSs from the Mayo Right 1K and Right 10K projects. The RIGHT 1K samples served as a control because we had already studied them individually. We found that 19 of 109 CYP2C9 and 36 of the 121 CYP2C19 missense variants that were studied displayed less than ~ 25% of the wild-type (WT) protein expression, a level that might have clinical relevance. We also compared variant calling by the DMS method with the predictions of computational algorithms and, finally, we validated severally damaging variants by the use of Western blot analysis. Our findings suggest that DMS can be an efficient method for the high-throughput identification of low protein abundance ORF variants that might have potential clinical implications.

**METHODS**

**Generation of landing pad HEK 293T cells**

The landing pad construct and the promoter-less cassette for CYP ORFs was created by Gibson assembly (Supplemental Text). TALENs were used to create double stranded breaks at the AAVS1 site and homology-directed repair was used to introduce landing pad constructs into HEK 293T cells. The landing pad construct expressed doxycycline inducible blue fluorescent protein (BFP), which was used as a selection marker for landing pad insertion. Thirty percent of HEK 293T cells display hypotriplot karyotypes. The assay requires that a single variant be integrated per cell. Single cells were sorted into each well of a 96-well plate for cloning. To identify which cells contained a single landing pad, we used Bxb1 recombinase mediated integration of a 1:1 ratio mixture of GFP and mCherry promoter-less plasmids and analyzed fluorescence by flow cytometry. Clones having the lowest percentage of both GFP and mCherry were most likely to be candidate for further assay. The attB-mCherry and attB-GFP plasmids were transfected 24 hours after Bxb1 recombinase transfection, and BFP induction by doxycycline. After 5 days, candidate clones were trypsinized, washed with phosphate-buffered saline, and were fixed in 4% formaldehyde at 4°C for 10 minutes. The cells were analyzed by flow cytometry.

**Nicking mutagenesis for variant library preparation**

Nicking mutagenesis methods were modified from Whitehead et al. to construct variant libraries for ORFs containing CYP2C9 and CYP2C19 missense variants. Nicking mutagenesis uses Nt.BbvCI and Nb.BbvCI sites and exonuclease cleavage to degrade WT template DNA. Nt.BbvCI enzyme with nicking endonuclease and exonuclease digestion of WT template DNA was used to form single stranded DNA. At the annealing step, five oligos carrying the variant of interest were annealed separately to single strand templates and then pooled with five reactions in one pot, using high fidelity DNA polymerase and Taq ligase to close the double strand. The second template strand was degraded by Nb.BbvCI endonuclease cleavage and exonuclease digestion, and a new second strand was synthesized.
with a common primer on a cassette plasmid backbone. Phosphorylated oligos for CYP2C9 and CYP2C19 variants were purchased from IDT (Coralville, IA). Sanger sequencing was used to validate the sequences of variant clones. Detailed protocols are provided in the Supplemental Text.

**Fluorescence-activated cell sorting**

Library cells were washed, trypsinized, and resuspended in phosphate-buffered saline containing 5% fetal bovine serum. Cells were then sorted on a FACSAria with 407, 488, and 532 nm lasers (BD Biosciences) into four bins, and the cells were collected in culture medium. GFP+/mCherry+ cells containing CYP2C9 or CYP2C19 variants were flow sorted and grown for 5 days. GFP+/mCherry+ cells were sorted again to determine the protein expression of CYP2C9/CYP2C19 variants based on their GFP/mCherry ratios. Gates were set based on GFP/mCherry ratios for cells integrating known CYP2C9/CYP2C19 variants and WT proteins as gating references. Four gates were set to dissect the pooled libraries into four different bins based on GFP/mCherry ratios. Data were analyzed by FACSDiva version 8.0.1 software.

**Variant calling**

Variant frequencies were calculated by high-throughput sequencing of the DNA collected in each sorted bin. Genomic DNA was extracted from sorted cells using DNA extraction kits (Qiagen) and amplicons were sequenced as described in the Supplemental Text. Fastq files were aligned with respective CYP2C9 and CYP2C19 reference sequences using BWA mem aligner version 0.7.15. Samtools mpileup version 1.5 was used with a custom python script for single nucleotide variation calling. VarScan pileup2indel version 2.3.9 was used for calling INDELS. A base quality score cutoff of 20 and a mapping quality score cutoff of 20 were applied for both single nucleotide variation and INDEL calling. Custom scripts were used to summarize the data and add allele frequencies at all positions in the reference sequence. Variant counts in each bin were tabulated and each variant’s frequency in each bin was calculated. The effects of variants were obtained based on the frequency of variants in each bin.

**RESULTS**

**Generation of landing pad cell lines**

The landing pad construct was generated and a promotor-less cassette was constructed with ORFs for CYP2C9 and CYP2C19. The ORFs were fused to GFP as an indicator of protein expression and mCherry, which was independently expressed as a transfection control. The DMS assay requires that only a single transgenetic cassette be integrated per cell, which could be guaranteed by having only a single landing pad per cell (Figure 1a). To generate a cell line that integrated only a single copy of the landing pad, the landing pad construct was inserted into HEK 293T cells and cells were selected by BFP. Single cells positive for BFP were sorted into each well of a 96-well plate for cloning. We observed different levels of BFP expression among different landing pad clones, probably because HEK 293T cells have been reported to have hypotriploid karyotypes. To identify a candidate clone with only a single landing pad, we selected >30 clones with low BFP expression levels. We then tested those clones by using bxb1 recombinase to integrate a mixture of 1:1 ratio GFP only and mCherry only promotor-less plasmids (plasmid sequences are shown in the Supplemental Text) into landing pad clones, followed by flow cytometry analysis. If a candidate clone contained only a single landing pad, it would integrate either GFP or mCherry, but not both. For clone 20, quadrant Q2 for the flow cytometry showed a negligible percentage of cells with both GFP and mCherry expression (Figure 1b). In addition, no overlay of green and red was observed in the fluorescence image for clone 20, once again indicating that clone 20 had only a single landing pad (Figure 1c). As a result, clone 20 was selected for use in the following experiments. This system allowed us to monitor variant protein expression in a high-throughput fashion.

**Effect of CYP2C9 and CYP2C19 variants on protein levels**

In our previous study, CYP2C9 218C>T, CYP2C9 343A>C, CYP2C19*3 (636G>A), and CYP2C19 815A>G affected final protein quantity, resulting in varying levels of decreased protein expression. In the present study, the reporter consisted of a promotor-less cassette containing the C-terminus of the CYP2C9 or CYP2C19 ORFs fused to GFP. That construct was expressed once it was integrated into landing pad cells to ensure one variant per cell and the landing pad BFP was disrupted. Flow cytometry analysis of GFP/mCherry+ cells showed that known “damaging” variants expressed significantly reduced levels of GFP and lower GFP/mCherry ratios, indicating that those cells expressed less protein. Mean GFP/mCherry ratios of known damaging variants were compared with those of WT proteins. The value for CYP2C9 218C>T was 28.4%; CYP2C9 343A>C was 48.7%; CYP2C19*3 (636G>A) was 17.6%; and CYP2C19 815A>G was 61.2% of the mean WT GFP/mCherry ratios, percentages that were roughly identical with Western blot results that we had previously reported (Figure 2a–c). Next, we created constructs for nonsynonymous variants with allele frequencies >0.0001 as reported by the Exome Aggregation Consortium and by the Mayo Clinic Right 1K study and created variant libraries for CYP2C9 and CYP2C19. Pooled variant libraries for both genes were integrated into landing pad cells. Using already known damaging variants for CYP2C9 and CYP2C19 as well as WT constructs as references for FACS gating, the cells were sorted into different “bins” based on GFP/mCherry ratios (Figure 2d). Variants of CYP2C9/CYP2C19 with the lowest GFP/mCherry ratios (<25% protein expression) were sorted into bin 1 and were classified as severely damaging. WT-like variants were sorted into bin 4. The gating for four-way sorting of CYP2C9/CYP2C19 pooled variant libraries is shown in Figure 3a,b. Pools of cells in each bin were collected and were then used as input material for DNA sequencing. We used NGS to monitor the frequencies of variants in each bin. Variant frequencies (Fv) appearing in each bin were then used to determine
protein abundance scores. Abundance scores for each CYP2C9 and CYP2C19 variant were calculated by use of the following equation:

\[
\text{Abundance score} = \frac{(F_{v,bin1} \times 0.25) + (F_{v,bin2} \times 0.5) + (F_{v,bin3} \times 0.75) + (F_{v,bin4} \times 1)}{(F_{v,bin1} + F_{v,bin2} + F_{v,bin3} + F_{v,bin4})}
\]

For each experiment, an “abundance score” for each variant studied was obtained by multiplying the variant frequency by weighted values (0.25–1) with bin1 assigned 0.25 and bin4 with 1.0.16 The final abundance score for each variant was calculated by averaging mean abundance scores across replicate assays. Variants were scored in at least three experiments, as shown graphically in Figure 4 and Figure S1. Variants were classified as “severely damaging,” “damaging,” or “tolerated,” with thresholds chosen on the basis of abundance scores for known CYP2C9/CYP2C19 variants.9 Using Western blot results and corresponding enzyme activities from our previous publication as a reference,9 CYP2C9 variants with abundance scores below 0.578 (CYP2C9 1076T>C), were classified as

**Figure 1** Generation of HEK 293T landing pad cells. (a) Plasmid maps of the landing pad construct and the promoter-less cassette for CYP open reading frames that were fused to green fluorescent protein (GFP) and engineered for the simultaneous expression of IRES-mCherry. (b) Flow cytometry results for landing pad HEK 293T clone 20 that had a low percentage of both GFP and mCherry, compatible with the conclusion that clone 20 contains a single landing pad. (c) Merged fluorescence photos showing that mCherry and GFP did not superimpose for clone 20. The landing pad clone 20 integrated either GFP or mCherry, which indicated that clone 20 contained a single landing pad. BFP, blue fluorescent protein; CMV, human cytomegalovirus promoter; IRES, internal ribosome entry site.
"severely damaging," displaying ~ 25% protein abundance as compared with WT, whereas those with abundance scores above that threshold but lower than 0.670 (CYP2C9 343A>C) were classified as "damaging" displaying ~ 50% of the protein abundance of the WT allozyme. Variants with scores above this threshold were considered "tolerated." CYP2C19 variants with abundance scores below 0.597 (CYP2C19 1349C>A), were classified as "severely damaging," whereas those with abundance scores above this threshold, but lower than 0.635 (CYP2C19 557G>A), were classified as "damaging." Variants with scores above this threshold were considered to be "tolerated." We found that 19 of 109 CYP2C9 and 36 of 121 CYP2C19 missense variants displayed less than ~ 25% of WT protein expression.

The variant calling results obtained by the use of DMS for variants from the ExAC study that had allele frequencies > 0.00001 and variants from the Mayo RIGHT 10K study are listed in Tables 1 and 2. The DMS results were also compared with predictions obtained using SIFT, Provean, and Polyphen2, and those results are listed in Tables S4 and S5 for CYP2C9 and CYP2C19, respectively. For variants that resulted in dramatically reduced protein expression levels, CYP2C9*11, CYP2C9*21, and another 12 CYP2C9 variants, as well as CYP2C19*8, CYP2C19*22, and another 26 CYP2C19 variants, the results were in good agreement among all three of these predictive algorithms. However, five CYP2C9 variants and seven CYP2C19 variants displayed < 25% of WT protein expression, whereas SIFT predicted that they were "tolerated." In summary, we found that the three commonly applied algorithms, which we tested on our 230 variants, disagreed among themselves 30.4% of the time, and they disagreed with our DMS assays, for at
least one algorithm, 54.7% of the time. We also searched PharmVar, a repository for pharmacogene variation that records functionally validated variants. For CYP2C9*21 and CYP2C19*2B, *8 and *22, our results were consistent with those reported in PharmVar. However, PharmVar currently lists several allozymes for both CYP2C9 and CYP2C19, which we found to be severely damaging as having only limited or moderate evidence. As a result, our results provide additional information with regard to the functional implications of these variants. Finally, 15 CYP2C9 and 30 CYP2C19 variants that we found to be severely damaging had not previously been reported or were reported to have uncertain function in PharmVar.

Validation of severely damaging variants of CYP2C9 and CYP2C19

*In silico* predictions were not always consistent with our DMS results, as described in the preceding paragraph. That fact emphasizes the need for the validation of variant classification using DMS or other functional assays. Although the efficiency of calling damaging variants using DMS exceeded the throughput of classical mutagenesis methods, we still needed to confirm the accuracy of calling for the variants that we studied. Therefore, we validated our variant protein expression calls by the use of Western blot analyses. As shown in Figure 5a,b, our newly identified severely damaging variants for CYP2C9 and CYP2C19 displayed significantly decreased protein expression (< 25% protein expression with the exception of CYP2C9 371G>T which had ~ 50% protein expression) compared with the WT protein—shown at the far right in each of the panels. The binning patterns for variant frequencies for selected allozyme are shown in Figures 5c,d.

**DISCUSSION**

The functional characterization of ORF missense variants in clinically important pharmacogenes remains a major challenge for pharmacogenomics. In a recent study, we identified and functionally characterized six novel nonsynonymous...
| Exact cDNA | Exact amino acid | RSID     | Common allele name | Allele frequency | Right 10K (variant prevalence) | Functional study | Abundance score | PharmVar |
|-----------|------------------|----------|--------------------|------------------|-------------------------------|-----------------|----------------|----------|
| c.752A>G  | p.His251Arg      | rs2256871| CYP2C9*9           | 0.006715         | 10079                         | 0.653893222     | No record      |
| c.449G>A  | p.Arg150His      | rs7900194| CYP2C9*8           | 0.005242         | 10074                         | 0.61743538      | Uncertain function |
| c.1003C>T | p.Arg335Trp      | rs28371685| CYP2C9*11          | 0.003784         | 10033                         | 0.549201141     | Possibly decreased |
| c.374G>A  | p.Arg125His      | rs7255819| CYP2C9*14          | 0.002953         | 10081                         | 0.627009829     | No record      |
| c.1465C>T | p.Pro489Ser      | rs9332239| CYP2C9*12          | 0.001912         | 10007                         | 0.583353141     | No record      |
| c.1080G>G | p.Asp360Glu      | rs28371666| CYP2C9*5           | 0.001163         | 10079                         | 0.629538267     | Moderate evidence |
| c.801C>T  | p.Phe267hPhe     | rs149158426|                 | 0.000855         | 10080                         | 0.597268637     | No record      |
| c.835G>A  | p.Pro279Thr      | rs182132424| CYP2C9*29          | 0.0040702        | 10072                         | 0.603761357     | Unknown function |
| c.1324G>A | p.Gly442Ser      | rs366545396|                 | 0.0002466        | 10077                         | 0.688673248     | No record      |
| c.1026A>G | p.Ser535Asp      | rs79001964| CYP2C9*11          | 0.0090851        | 10033                         | 0.639522141     | Unknown function |
| c.1264A>G | p.Ser422Gly      | rs77679484| CYP2C9*14          | 0.0020247         | 10082                         | 0.627667871     | No record      |
| c.895A>G  | p.Thr299Ala      | rs72558192| CYP2C9*16          | 0.0030247         | 10079                         | 0.673073074     | No record      |
| c.520C>G  | p.Pro174Ala      | rs199539783|                 | 0.001649         | 10079                         | 0.651742423     | No record      |
| c.296N>C  | p.Leu90Pro       | rs72558187| CYP2C9*13          | 0.001401         | 10079                         | 0.616189783     | Possibly decreased |
| c.431G>A  | p.Arg144His      | rs14189852|                 | 0.001401         | 10079                         | 0.651784299     | No record      |
| c.1084G>G | p.Leu362Val      | rs150663116|                 | 0.000132         | 10082                         | 0.581270896     | Possibly decreased |
| c.1088C>T | p.Pro363Leu      | rs15066316|                 | 0.000139         | 10074                         | 0.689018975     | No record      |
| c.449G>T  | p.Arg150Leu      | rs149158426|                 | 0.0002136        | 10079                         | 0.608683972     | No record      |
| c.515G>C  | p.Cys172Ser      | rs14761799|                 | 0.0001155        | 10079                         | 0.748597212     | No record      |
| c.1034T>C | p.Met345Thr      | rs72558192|                 | 0.0001072        | 10079                         | 0.601859628     | No record      |
| c.343A>G  | p.Ser115Arg      | rs139532088|                 | 0.0009894        | 10082                         | 0.57862159      | Uncertain function |
| c.980C>T  | p.Ile327Thr      | rs75507570| CYP2C9*31          | 0.0009007        | 10082                         | 0.611262013     | No record      |
| c.89C>T   | p.Pro30Leu       | rs142240658| CYP2C9*21          | 0.0009067        | 10082                         | 0.483953225     | Uncertain function |
| c.418C>T  | p.Arg150Cys      | rs1784037| CYP2C9*14          | 0.0009066        | 10082                         | 0.710010892     | No record      |
| c.688G>C  | p.His230Asp      | rs149158426|                 | 0.000837         | 10080                         | 0.65625         | Unknown function |
| c.389C>T  | p.Ala130Met      | rs2009695026| CYP2C9*44          | 0.0008248        | 10080                         | 0.639755882     | No record      |
| c.17G>A   | p.Ala173Thr      | rs373758966|                 | 0.0008247        | 10080                         | 0.548579579     | No record      |
| c.395C>A  | p.Arg132Gln      | rs200183364| CYP2C9*33          | 0.0008246        | 10083                         | 0.711347628     | No record      |
| c.1362G>C | p.Glu454His      | rs72558192| CYP2C9*19          | 0.0008243        | 10082                         | 0.645897332     | No record      |
| c.800C>T  | p.Pro227Leu      | rs530905257|                 | 0.000756         | 10082                         | 0.705389146     | Possibly decreased |
| c.143C>T  | p.Pro480Leu      | rs530905257|                 | 0.0007417        | 10082                         | 0.479818446     | Possibly decreased |
| c.556C>T  | p.Arg186Cys      | rs150435881|                 | 0.0006602        | 10082                         | 0.592060606     | Possibly decreased |

(Continues)
| Exact cDNA | Exact amino acid | RSID   | Common allele name | Allele frequency | Right 10K (variant prevalence) | DMS          | Functional study | Abundance score |
|------------|-----------------|--------|--------------------|------------------|-------------------------------|--------------|-----------------|------------------|
| c.116T>C   | p.Ile389Thr     |        |                    | 0.00006957       | 10082 2 0                     | Damaging     | 0.631849327     | No record        |
| c.1421A>G  | p.Asn474Ser     | rs141011391 |                    | 0.00006953       | 10083 1 0                     | Severe Damaging | 0.543305795     | uncertain function |
| c.373G>T   | p.Arg125Cys     |        |                    | 0.00005775       | 10081 3 0                     | TOLERATED    | 0.677726163     | no record        |
| c.1187A>G  | p.His396Arg     | rs187793133 |                    | 0.00005772       | 10081 3 0                     | TOLERATED    | 0.700935174     | no record        |
| c.1370A>G  | p.Asn457Ser     |        |                    | 0.00005777       | 10081 3 0                     | TOLERATED    | 0.676032437     | no record        |
| c.263T>C   | p.Ile88Thr      | rs139656048 | CYP2C9*61          | 0.00005768       | 10083 1 0                     | Severe Damaging | 0.571077996     | uncertain function |
| c.218G>T   | p.Pro73Leu      | rs762081829 |                    | 0.00005699       | 10083 1 0                     | TOLERATED    | 0.656125305     | no record        |
| c.595G>A   | p.Glu199Lys     |        |                    | 0.00004956       | 10083 1 0                     | Damaging     | 0.58402074      | uncertain function |
| c.371G>A   | p.Arg124Gln     | rs12444460 | CYP2C9*42          | 0.00004945       | 10083 1 0                     | Damaging     | 0.656819836     | uncertain function |
| c.458T>C   | p.Val153Ala     | rs37399395 |                    | 0.00004934       | 10083 1 0                     | TOLERATED    | 0.654017348     | no record        |
| c.719T>C   | p.Met240Thr     |        |                    | 0.00004046       | 10083 1 0                     | Damaging     | 0.58402074      | uncertain function |
| c.619A>T   | p.Ile207Phe     |        |                    | 0.00004034       | 10083 1 0                     | TOLERATED    | 0.674486125     | no record        |
| c.371G>T   | p.Arg124Leu     | rs12444460 | CYP2C9*34          | 0.00004024       | 10083 1 0                     | Severe Damaging | 0.568764708     | uncertain function |
| c.164G>T   | p.Thr55Ile      | rs771905380|                    | 0.00004014       | 10083 1 0                     | TOLERATED    | 0.701629687     | no record        |
| c.1004G>A  | p.Arg335Gln     |        | CYP2C9*42          | 0.00004004       | 10083 1 0                     | TOLERATED    | 0.695204529     | no record        |
| c.709G>A   | p.Val237Ile     |        |                    | 0.00003994       | 10083 1 0                     | Severe Damaging | 0.529048923     | uncertain function |
| c.624G>C   | p.Leu208Phe     |        |                    | 0.00003984       | 10083 1 0                     | TOLERATED    | 0.681854013     | no record        |
| c.538T>C   | p.Ser180Pro     |        |                    | 0.00003974       | 10083 1 0                     | Severe Damaging | 0.529048923     | uncertain function |
| c.541A>G   | p.Ile181Val     |        |                    | 0.00003964       | 10083 1 0                     | Severe Damaging | 0.529048923     | no record        |
| c.1297C>T  | p.Arg433Trp     |        |                    | 0.00003954       | 10083 1 0                     | TOLERATED    | 0.689873007     | no record        |
| c.386T>A   | p.Met129Lys     | rs750097042 |                    | 0.00003944       | 10083 1 0                     | Damaging     | 0.634131636     | no record        |
| c.122A>T   | p.Asn41Ile      |        |                    | 0.00003934       | 10083 1 0                     | Severe Damaging | 0.458665155     | no record        |
| c.1024A>G  | p.Arg342Gly     |        |                    | 0.00003924       | 10083 1 0                     | Severe Damaging | 0.575108971     | normal function  |
| c.1036G>T  | p.Pro346Ser     |        |                    | 0.00003914       | 10083 1 0                     | Damaging     | 0.654218921     | no record        |
| c.1429G>A  | p.Ala477Thr     |        | CYP2C9*30          | 0.00003904       | 10083 1 0                     | TOLERATED    | 0.692533067     | no record        |
| c.632C>T   | p.Pro211Leu     |        |                    | 0.00003894       | 10083 1 0                     | TOLERATED    | 0.64966918      | no record        |
| c.629G>A   | p.Ser210Asn     |        |                    | 0.00003884       | 10083 1 0                     | TOLERATED    | 0.699592318     | no record        |
| c.600G>T   | p.Lys200Asn     | rs766903671 |                    | 0.00003874       | 10083 1 0                     | Damaging     | 0.663003421     | no record        |
| c.700G>T   | p.Arg124Gln     |        | CYP2C9*43          | 0.00003864       | 10083 1 0                     | TOLERATED    | 0.737650497     | no record        |
| c.1153A>T  | p.Thr385Ser     |        |                    | 0.00003854       | 10083 1 0                     | TOLERATED    | 0.601528578     | no record        |
| c.863A>G   | p.Glu288Gly     |        |                    | 0.00003844       | 10083 1 0                     | Damaging     | 0.66603031      | no record        |
| c.1135T>C  | p.Tyr379His     |        |                    | 0.00003834       | 10083 1 0                     | Damaging     | 0.66603031      | no record        |
| c.433G>T   | p.Val145Phe     |        |                    | 0.00003824       | 10083 1 0                     | Severe Damaging | 0.460289634     | uncertain function |
| c.1022A>G  | p.Asp341Gly     |        |                    | 0.00003814       | 10083 1 0                     | Severe Damaging | 0.470510412     | no record        |
| c.704G>A   | p.Lys225Arg     |        |                    | 0.00003804       | 10083 1 0                     | TOLERATED    | 0.68009768      | uncertain function |
| c.791T>C   | p.Ile264Ser     | rs761995497 |                    | 0.00003794       | 10083 1 0                     | Damaging     | 0.599331078     | uncertain function |
| c.773A>G   | p.Asn258Ser     |        |                    | 0.00003784       | 10083 1 0                     | TOLERATED    | 0.684042216     | no record        |
| Exact cDNA | Exact amino acid | RSID   | Common allele name | Allele frequency | Right 10K (variant prevalence) | DMS |
|-----------|------------------|--------|--------------------|------------------|-------------------------------|-----|
| c.638T>C  | p.Ile213Thr      | .      |                    | 0.00001655       | Damaging                      | 0.648500178 | No record |
| c.637A>C  | p.Ile213Leu      | .      |                    | 0.00001655       | TOLERATED                     | 0.676605037 | No record |
| c.618G>T  | p.Lys206Asn      | .      |                    | 0.00001654       | TOLERATED                     | 0.690691147 | No record |
| c.356A>C  | p.Lys119Thr      | .      | CYP2C9*41          | 0.00001651       | Damaging                      | 0.601033475 | decreased function |
| c.1076T>C | p.Ile359Thr      | rs56165452 | CYP2C9*4 | 0.0000165 | Severely Damaging | 0.578003356 | Possibly decreased |
| c.389C>A  | p.Thr130Lys      | .      |                    | 0.00001656       | Damaging                      | 0.668748634 | Unknown function |
| c.820G>A  | p.Glu274Lys      | .      |                    | 0.00001657       | TOLERATED                     | 0.678495665 | No record |
| c.1080C>A | p.Asp360Glu      | rs28371686 | CYP2C9*4 | 0.0000165 | Damaging                      | 0.605882567 | No record |
| c.137G>A  | p.Gly46Asp       | .      |                    | 0.00001649       | TOLERATED                     | 0.611297817 | No record |
| c.109C>T  | p.Pro37Ser       | .      |                    | 0.00001649       | Damaging                      | 0.62487462 | No record |
| c.1045G>A | p.Asp49Asn       | .      |                    | 0.00001649       | Damaging                      | 0.648954995 | No record |
| c.908G>T  | p.Ser303Ile      | .      |                    | 0.00001649       | Damaging                      | 0.653594136 | No record |
| c.986G>A  | p.Arg329His      | .      |                    | 0.00001649       | Damaging                      | 0.66037359 | Unknown function |
| c.1136A>G | p.Tyr379Gly      | rs773074286 | CYP2C9*4 | 0.00001649 | Damaging                      | 0.665754704 | No record |
| c.1214A>T | p.Glu405Val      | .      |                    | 0.00001649       | TOLERATED                     | 0.675922464 | No record |
| c.146A>G  | p.Asp49Gly       | .      | CYP2C9*37          | 0.00001649       | TOLERATED                     | 0.678200758 | Uncertain function |
| c.422T>C  | p.Ile141Thr      | rs146815754 | CYP2C9*46 | 0.00001649 | Damaging                      | 0.681659253 | No record |
| c.1159A>G | p.Ile387Val      | rs764211126 | CYP2C9*46 | 0.00001649 | Damaging                      | 0.688530952 | No record |
| c.968T>C  | p.Val323Ala      | .      |                    | 0.00001649       | Damaging                      | 0.700837059 | No record |
| c.989G>C  | p.Val303Ala      | .      |                    | 0.00001649       | Damaging                      | 0.727107621 | No record |
| c.257G>A  | p.Ala86Asp       | .      |                    | 0.00001648       | Damaging                      | 0.513086248 | No record |
| c.388G>A  | p.Gly149Val      | .      | CYP2C9*46          | 0.00001648       | Damaging                      | 0.590756013 | Uncertain function |
| c.445G>A  | p.Ala149Thr      | .      | CYP2C9*46          | 0.00001648       | Damaging                      | 0.604943716 | No record |
| c.312A>T  | p.Glu104Asp      | .      |                    | 0.00001648       | Damaging                      | 0.638095238 | No record |
| c.271G>A  | p.Gly91Arg       | .      |                    | 0.00001648       | Damaging                      | 0.651227774 | No record |
| c.296T>A  | p.Ile99Asn       | .      |                    | 0.00001648       | Damaging                      | 0.651227774 | No record |
| c.1415T>C | p.Val472Ala      | .      |                    | 0.00001648       | Damaging                      | 0.668443691 | No record |
| c.460G>A  | p.Glu154Lys      | .      |                    | 0.00001648       | Damaging                      | 0.669561409 | Uncertain function |
| c.296T>C  | p.Ile99Thr       | .      |                    | 0.00001648       | Damaging                      | 0.68904038  | Decreased function |
| c.8C>A   | p.Ser3Tyr        | .      |                    | 0.00001648       | Damaging                      | 0.6911887 | No record |
| c.791T>C  | p.Ile264Thr      | rs761895497 | CYP2C9*46 | 0.0000109 | Damaging                      | 0.683923331 | Unknown function |
| c._del707 | p.Asn236Thr*5    | .      |                    | 0.000004096 | Damaging                      | 0.623703953 | No record |
| c.709G>C  | p.Val237Leu      | .      |                    | Not recorded      | TOLERATED                     | 0.72516812 | No record |
| c.229G>T  | p.Leu77Met       | .      |                    | Not recorded      | TOLERATED                     | 0.72516812 | No record |

DMS, deep mutational scanning; ExAC, Exome Aggregation Consortium; RSID, Reference SNP ID number; WT, wild-type.
| Exact cDNA | Exact Amino acid | RSID | Common allele name | Allele Frequency | Right 10K (variant prevalence) | DMS | Functional study | Abundance score | PharmVar |
|-----------|-----------------|------|--------------------|------------------|-----------------------------|-----|-----------------|----------------|----------|
| c.919G>A  | V331I           | rs3758581 |                    | 0.06242          | 8802                        |     | Damage          | 0.599924823 | Normal function |
| c.276G>C  | E92D            | rs17878459 | CYP2C19*2B         | 0.0236           | 9480                        |     | TOLERATED       | 0.713318402 | No record |
| c.518G>T  | A173V           | rs61311738 |                    | 0.005818         | 10080                       |     | TOLERATED       | 0.643519769 | No record |
| c.481G>T  | A161P           | rs181297724 |                 | 0.004786         | 10080                       |     | TOLERATED       | 0.713318402 | No record |
| c.449G>A  | R150H           | rs58973490 | CYP2C19*11         | 0.002652         | 9985                        |     | TOLERATED       | 0.692714941 | normal function |
| c.55A>C   | I19L            | rs17882687 | CYP2C19*15         | 0.002127         | 10079                       |     | Damaging        | 0.623042396 | normal function |
| c.358T>C  | W120R           | rs41291556 | CYP2C19*8          | 0.00154          | 10045                       |     | Damaging        | 0.47878285 | No function |
| c.1228C>T | R410C           | rs17879685 | CYP2C19*13         | 0.001517         | 10082                       |     | TOLERATED       | 0.676550868 | Normal function |
| c.431G>A  | R144H           | rs17884712 | CYP2C19*9          | 0.001079         | 10080                       |     | Damaging        | 0.621114695 | decreased function |
| c.365A>C  | E122A           | rs17885179 |                    | 0.0009637        | 10082                       |     | TOLERATED       | 0.753787879 | No record |
| c.784G>A  | D262N           | .      |                    | 0.0008586        | 10082                       |     | Damaging        | 0.614038591 | No record |
| c.448C>T  | R150C           | rs14297481 |                    | 0.0004859        | 30                          |     | Damaging        | 0.62953438 | No record |
| c.557G>A  | R186C           | rs145328984 |                | 0.000448         | 10079                       |     | Damaging        | 0.62712679 | decreased function |
| c.985C>T  | R329C           | rs59734894 |                    | 0.0003872        | 10078                       |     | Severely damaging | 0.554502254 | No record |
| c.394C>T  | R132W           | rs149590953 |                 | 0.0003871        | 10081                       |     | Damaging        | 0.635242943 | No record |
| c.241G>A  | E81K            | rs14902229 |                    | 0.0003871        | 10068                       |     | Tolerated       | 0.66776879 | No record |
| c.374G>A  | R125H           | rs141774245 |                | 0.0002965        | 10062                       |     | Tolerated       | 0.615675954 | No function |
| c.337G>A  | V131I           | rs145119820 |                | 0.0002718        | 10079                       |     | Tolerated       | 0.68388581 | No record |
| c.129A>T  | K432I           | rs146991374 |                | 0.000264         | 10082                       |     | Damaging        | 0.6082438 | No record |
| c.217C>T  | R73C            | rs145328984 | CYP2C19*30        | 0.0002553         | 10082                       |     | Damaging        | 0.623329263 | uncertain function |
| c.1078G>A | D360N           | rs144036596 |                | 0.0002471        | 10082                       |     | TOLERATED       | 0.693150526 | None |
| c.395C>A  | R132Q           | rs72552267 | CYP2C19*6         | 0.0002389         | 10082                       |     | TOLERATED       | 0.718026114 | No function |
| c.164G>G  | T55S            | rs572853437 |                | 0.000132         | 10082                       |     | TOLERATED       | 0.718673647 | No record |
| c.557G>C  | R186P           | rs140278421 | CYP2C19*22        | 0.0001236         | 10083                       |     | Tolerated       | 0.560050139 | No function |
| c.1004G>A | R335Q           | rs118203757 | CYP2C19*24        | 0.0001236         | 10080                       |     | Tolerated       | 0.601312932 | no function |
| c.557G>A  | R186H           | .      |                    | 0.00108          | 10080                       |     | Tolerated       | 0.653198497 | No record |
| c.831C>A  | N277K           | .      |                    | 9.888E-05        | 10080                       |     | Tolerated       | 0.681062109 | No record |
| c.1048G>A | A350T           | rs201509150 |                | 9.884E-05        | 10080                       |     | Severely damaging | 0.536022841 | No record |
| c.1127T>A | F376Y           | .      |                    | 9.884E-05        | 10083                       |     | Tolerated       | 0.711485306 | No record |
| c.25C>G   | L9V             | .      |                    | 9.884E-05        | 10083                       |     | Tolerated       | 0.731784659 | No record |
| c.1007G>T | S336I           | rs143833145 |                | 9.061E-05        | 10080                       |     | Tolerated       | 0.63737252 | No record |
| c.1036C>T | P346S           | .      |                    | 9.000906         | 10080                       |     | Tolerated       | 0.80346079 | No record |
| c.556C>T  | R186C           | rs183701923 |                | 8.242E-05        | 10082                       |     | Severely damaging | 0.564313994 | No record |
| c.389C>T  | T130M           | rs150152656 |                | 8.237E-05        | 10082                       |     | TOLERATED       | 0.684802678 | No record |

(Continues)
Table 2  (Continued)

| Exact cDNA | Exact Amino acid | Common allele name | Allele Frequency | Right 10K (variant prevalence) | DMS |
|------------|------------------|--------------------|------------------|---------------------------------|-----|
| c.527A>G   | N176S            | rs57700608         | 7.417E-05        | 10074 10 0                      | TOLERATED 0.671875 No record |
| c.1075A>G  | I359V            | .                  | 7.414E-05        |                                 | Damaging 0.630364833 No record |
| c.440A>G   | E147G            | rs147453531        | 7.413E-05        | 10081 3 0                      | TOLERATED 0.63455877 No record |
| c.781C>T   | R261W            | .                  | 6.605E-05        |                                 | Severe damaging 0.501301905 No record |
| c.836A>C   | Q279P            | rs615263399        | 6.591E-05        |                                 | Severe damaging 0.57090791 No record |
| c.1001A>T  | N334I            | .                  | 0.0000659        |                                 | Severe damaging 0.648759438 No record |
| c.221T>C   | M74T             | .                  | 6.589E-05        |                                 | No record |
| c.1324C>T  | R442C            | rs192154563        | 5.772E-05        |                                 | No record |
| c.1021G>C  | D341H            | rs770829708        | 5.766E-05        | 10081 3 0                      | TOLERATED 0.656054948 No record |
| c.85C>T    | L29F             | .                  | 4.946E-05        |                                 | TOLERATED 0.692408379 No record |
| c.837G>T   | Q279H            | .                  | 4.944E-05        |                                 | TOLERATED 0.678912244 No record |
| c.1003C>T  | R335W            | rs368758960        | 4.942E-05        | 10083 1 0                      | Severe damaging 0.497372294 No record |
| c.1034T>A  | M345K            | rs201132803        | 4.942E-05        | 10080 1 0                      | Severe damaging 0.497916398 No record |
| c.371G>A   | R124Q            | rs200346442        | 4.942E-05        |                                 | Damaging 0.602183853 No record |
| c.925G>A   | A309T            | .                  | 4.942E-05        |                                 | TOLERATED 0.638680409 No record |
| c.218G>A   | R73H             | .                  | 4.119E-05        |                                 | Severe damaging 0.476206097 No record |
| c.1072C>T  | Y358H            | .                  | 4.119E-05        |                                 | TOLERATED 0.648167478 No record |
| c.370C>T   | R124W            | .                  | 4.118E-05        |                                 | Severe damaging 0.505962416 No record |
| c.726T>A   | S242R            | .                  | 3.311E-05        |                                 | TOLERATED 0.702093879 No record |
| c.801C>A   | F267L            | rs377674118        | 3.303E-05        | 10083 1 0                      | Severe damaging 0.537597325 No record |
| c.1171C>G  | L391V            | .                  | 3.298E-05        |                                 | Severe damaging 0.569318436 No record |
| c.1205C>T  | P402L            | .                  | 3.298E-05        |                                 | TOLERATED 0.642305732 No record |
| c.1216A>G  | M406V            | rs144056033        | 3.298E-05        | 10082 2 0                      | TOLERATED 0.699731714 No record |
| c.1060G>A  | E354K            | .                  | 3.295E-05        |                                 | Severe damaging 0.475659325 No record |
| c.896C>G   | T299R            | .                  | 3.295E-05        |                                 | Severe damaging 0.531790263 No record |
| c.305T>C   | L102P            | .                  | 3.295E-05        |                                 | Severe damaging 0.574479079 No record |
| c.905C>T   | T302l            | rs58259047         | 3.295E-05        |                                 | Damaging 0.602678118 No record |
| c.1120G>A  | V374I            | rs113934938        | 3.295E-05        |                                 | Damaging 0.627074235 normal function |
| c.961G>T   | A321S            | .                  | 3.295E-05        |                                 | TOLERATED 0.639949178 No record |
| c.928C>T   | L310F            | .                  | 3.295E-05        | 10082 2 0                      | TOLERATED 0.679301841 No record |
| c.907A>G   | S303G            | .                  | 3.295E-05        |                                 | TOLERATED 0.717662336 No record |
| c.648C>G   | C216W            | .                  | 0.0000274        |                                 | TOLERATED 0.856662248 No record |
| c.671A>T   | D224T            | .                  | 2.561E-05        |                                 | Severe damaging 0.538999792 No record |
| c.753C>G   | H251Q            | rs148247410        | 2.478E-05        |                                 | Severe damaging 0.564668153 No function |
| c.769A>T   | I257F            | .                  | 2.477E-05        |                                 | TOLERATED 0.681306328 No record |
| c.631C>A   | P211T            | .                  | 2.476E-05        |                                 | TOLERATED 0.659306032 No record |

(Continues)
| Exact cDNA | Exact Amino acid | RSID       | Common allele name | Allele Frequency | Right 10K (variant prevalence) | Functional study | Abundance score | PharmVar |
|-----------|-----------------|------------|-------------------|-----------------|-------------------------------|-----------------|----------------|----------|
| c.629C>A  | T210N           | rs370803989| CYP2C19*35        | 2.473E-05       | 10080            | TOLERATED                  | 0.67797182      | uncertain |
| c.82A>G   | K28E            |            |                   | 2.473E-05       | TOLERATED                  | 0.680943598     | No record      |          |
| c.185G>C  | G62A            |            |                   | 2.472E-05       | TOLERATED                  | 0.650769501     | No record      |          |
| c.190G>A  | V64M            | rs150045105|                   | 2.472E-05       | TOLERATED                  | 0.65615054      | No record      |          |
| c.848C>T  | T283I           |            |                   | 2.472E-05       | Damage                      | 0.619158302     | No record      |          |
| c.1037C>T | P346L           |            |                   | 2.471E-05       | TOLERATED                  | 0.68887841      | No record      |          |
| c.1080C>A | D360E           |            |                   | 2.471E-05       | Severely damaging          | 0.545128622     | No record      |          |
| c.1078G>C | D360H           | rs144036596|                   | 2.471E-05       | Severely damaging          | 0.567361111     | No record      |          |
| c.373C>T  | R125C           | rs200150287|                   | 2.471E-05       | Severely damaging          | 0.669159641     | No record      |          |
| c.850A>G  | I284V           |            |                   | 2.471E-05       | Severely damaging          | 0.685908298     | No record      |          |
| c.430C>T  | R144C           |            |                   | 2.471E-05       | Severely damaging          | 0.68887841      | No record      |          |
| c.721G>A  | E241K           |            |                   | 1.655E-05       | Severely damaging          | 0.577527583     | No record      |          |
| c.1288G>C | A430P           |            |                   | 1.655E-05       | Severely damaging          | 0.577527583     | No record      |          |
| c.813G>A  | M271I           |            |                   | 1.655E-05       | Severely damaging          | 0.577527583     | No record      |          |
| c.169C>G  | L57V            |            |                   | 1.655E-05       | Severely damaging          | 0.577527583     | No record      |          |
| c.1465C>T | P489S           |            |                   | 0.0000165       | Severely damaging          | 0.577527583     | No record      |          |
| c.137T>T  | L458R           | rs150187034 |                   | 1.649E-05       | Severely damaging          | 0.51802352      | No record      |          |
| c.1349C>A | T450N           | rs141690375 |                   | 1.649E-05       | Severely damaging          | 0.597025262     | No record      |          |
| c.1439C>T | P480L           | rs177950172 |                   | 1.649E-05       | Severely damaging          | 0.626171339     | No record      |          |
| c.1325G>A | R442H           | rs138112316 |                   | 1.649E-05       | Severely damaging          | 0.646762738     | No record      |          |
| c.65A>G   | Q22R            | rs144982727 |                   | 1.649E-05       | Severely damaging          | 0.656465743     | No record      |          |
| c.1398C>A | D466E           |            |                   | 1.649E-05       | Severely damaging          | 0.661470643     | No record      |          |
| c.593T>C  | M198T           | rs186489608 |                   | 1.649E-05       | Severely damaging          | 0.667063402     | No record      |          |
| c.1330G>C | E444Q           |            |                   | 1.649E-05       | Severely damaging          | 0.694466074     | No record      |          |
| c.1213G>C | E405Q           |            |                   | 1.649E-05       | Severely damaging          | 0.711027827     | No record      |          |
| c.518C>A  | A173D           | rs61311738  |                   | 1.648E-05       | Severely damaging          | 0.49614976      | No record      |          |
| c.1076T>A | G59N            |            |                   | 1.648E-05       | Severely damaging          | 0.52206232      | No record      |          |
| c.197G>T  | T66S            |            |                   | 1.648E-05       | Severely damaging          | 0.52206232      | No record      |          |
| c.862G>A  | V288I           |            |                   | 1.648E-05       | Severely damaging          | 0.673305956     | No record      |          |
| c.537C>G  | C179W           |            |                   | 1.648E-05       | Severely damaging          | 0.694502569     | No record      |          |
| c.865A>G  | I289V           |            |                   | 1.648E-05       | Severely damaging          | 0.714496776     | No record      |          |
Table 2 (Continued)

| Allele name | Frequency | Abundance | Functional study | ORF variant function | Reference SNP ID number | WT, wild-type |
|-------------|-----------|-----------|------------------|----------------------|-------------------------|---------------|
| c.445G>A    | 1.647E-05 | 1         | Severely damaging| TOLERATED            | rs22908787              | 0.521050873  |
| c.905C>G    | 1.647E-05 | 1         | Severely damaging| TOLERATED            | rs58259047              | 0.521776385 |
| c.331G>A    | 1.647E-05 | 1         | Severely damaging| TOLERATED            | rs22908787              | 0.541454278 |
| c.1021G>A   | 1.647E-05 | 1         | Severely damaging| TOLERATED            | rs22908787              | 0.571470548 |
| c.409G>C    | 1.647E-05 | 1         | Severely damaging| TOLERATED            | rs22908787              | 0.571889977 |
| c.1013G>T   | 1.647E-05 | 1         | Damaging         | TOLERATED            | rs22908787              | 0.609881164 |
| c.217C>A    | 10082     | 1         | TOLERATED        | TOLERATED            | rs145328984             | 0.651406999 |
| c.410G>A    | 1.647E-05 | 1         | TOLERATED        | TOLERATED            | rs22908787              | 0.653990969 |
| c.326G>C    | 1.647E-05 | 1         | TOLERATED        | TOLERATED            | rs22908787              | 0.658559311 |
| c.347A>G    | 1.647E-05 | 1         | TOLERATED        | TOLERATED            | rs22908787              | 0.666187952 |
| c.271G>C    | 1.647E-05 | 1         | TOLERATED        | TOLERATED            | rs22908787              | 0.668206476 |
| c.1002C>A   | 1.647E-05 | 1         | TOLERATED        | TOLERATED            | rs563052490             | 0.687674269 |
| c.1112C>G   | 1.647E-05 | 1         | TOLERATED        | TOLERATED            | rs568155950             | 0.715706169 |
| c.578A>G    | 4.06E-06  | 1         | TOLERATED        | TOLERATED            | rs22908787              | 0.658206476 |

ExAC, Exome Aggregation Consortium; RSID, Reference SNP ID number; WT, wild-type.

Data from Table 2 (Continued) shows the variants identified in the CYP2C9 and CYP2C19 genes. The table includes the variant name, frequency, and functional study results for each variant. The table also indicates whether the variant is homozygous or heterozygous and the abundance score.

ORF variants in CYP2C9 and seven in CYP2C19 based on Mayo Right 1K data. We found that the enzyme activities of these variants generally correlated well with protein expression levels. Missense variants in CYP2C9/CYP2C19 may alter protein folding, leading to decreased protein expression as a result of accelerated proteasome-mediated degradation, a major factor responsible for decreased enzymatic activity in pharmacogenomics. The loss of function of allozymes containing nonsynonymous CYP2C9/CYP2C19 ORF single nucleotide polymorphisms due to decreased protein expression made it possible to analyze their function by fluorescence reporter assays. Because of the very large number of missense variants in ORFs, it is practically difficult to link the genotypes of these variants to their functional phenotypes using "one-at-a-time" expression systems. Fowler and colleagues developed DMS to analyze variants for all possible amino acid alterations in several genes. Saturation mutagenesis was also used for DMS in several previous studies. That approach has advantages for use in pre-emptive pharmacogenomics and makes it possible to interpret variant function based on protein structural mapping. Degenerate codons were used to generate the saturation libraries but some variants of interest may be missed due to codon bias, with up to 30–50% of possible variants missing from the final data sets. CYP2C9 and CYP2C19 each have ORFs that are 1.6 kb in length, so it would be difficult to generate mutation libraries for all possible amino acid alterations in several genes. Saturation mutagenesis was used for DMS in several previous studies. That approach has advantages for use in pre-emptive pharmacogenomics and makes it possible to interpret variant function based on protein structural mapping.
validate results obtained by using predictive algorithms. \(^{30,31}\) Therefore, we compared variant calling by use of DMS with the predictions of computational algorithms, and differences were found between our results and those of prediction algorithms, as listed in Tables S4 and S5. Those differences may be due to either the accuracy of the prediction algorithms or to underlying molecular mechanisms. For example, CYP2C9 709G>C and CYP2C19 65A>G and CYP2C19*13 have WT-like protein expression but loss of enzyme activity.\(^7\)\(^,\)\(^32\) We also determined whether variants identified by DMS were truly damaging by the use of Western blot analyses, as shown in Figure 5.

A limitation of DMS based on fluorescence is that some genes have been found to not be amenable to this assay, because damaging variants for those genes had similar fluorescent intensities as did WT proteins.\(^{16}\) We found that DMS seems to be most sensitive for screening for severely damaging variants, which displayed clear fluorescence separation.
from WT-like variants. However, it required careful interpretation of intermediate-fluorescing variants. The validation of functional studies for variants characterized in this fashion will be essential if we are to incorporate these results into clinical decision making and electronic health records. To validate the most severely damaging variants that we identified by DMS, we used Western blot assays as our standard functional assay for validation, even though those studies were laborious and time-consuming—but still necessary at this time. Protein expression is obviously an important aspect of functional genomics—but only one aspect. Additional functional validation based on enzyme activities for different CYP substrates should be performed in the future to further extend the functional characterization of the variant alleles that we have studied. The regulation of CYP activity is a complex process involving multiple mechanisms, which include transcripton regulation by nuclear receptors, such as the pregnane X receptor, the constitutive androstane receptor, the glucocorticoid receptor and by members of the TSPYL gene family.23-38 DMS, as we have used it, has the limitation of not addressing upstream DNA sequence variants, such as CYP2C19*17 that results in an increase in transcriptional activity.37-39 High-throughput methodology for studying variants outside of ORFs will obviously be required for the interpretation of CYP variants that map outside of protein coding sequences. As a result, DMS is not a “final answer” but rather represents a significant step forward in our efforts to link genomic variation to variation in drug response phenotypes.

In summary, we have identified and validated 15 CYP2C9 and 30 CYP2C19 severely damaging variants that had not previously been reported in PharmVar.20 Those variants are potentially clinically actionable. Functional studies of those variants showed decreased protein expression, which could result in decreased drug metabolism. Our results add information that may help to improve the accuracy of current prediction algorithms and they may also provide test data sets for machine-learning methods that might “learn” to predict the effects of ORF missense variants. The Mayo Clinic recently expanded the RIGHT 1K study to include an even larger cohort, the RIGHT 10K study that includes an additional 10,085 DNA samples with sequencing data for 77 pharmacogenes.9 Single nucleotide polymorphisms from both RIGHT 1K and RIGHT 10K studies were included in our analyses. This same DMS methodology can now be implemented to study other important pharmacogenes and preemptive NGS Mayo RIGHT 10K data as ever larger numbers of ORF missense variants are identified.

Supporting Information. Supplementary information accompanies this paper on the Clinical and Translational Science website (www.cts-journal.com).

Figure S1.  
Table S1.  
Table S2.  
Table S3.  
Table S4.  
Table S5.  
Table S6.  
Supplemental Text.

Funding. This study was funded by National Institutes of Health (NIH) grants U19 GM61388 (The Pharmacogenomics Research Network), R01 GM28157, R01 GM125633, and T32 GM08685, and by the Mayo Clinic Center for Individualized Medicine.

Conflict of Interest. Both Drs. Weinshilboum and Wang are co-founders of and stockholders in Oneome, LLC. All other authors declared no competing interests for this work.

Author Contributions. L.Z., V.S., J.Y., D.L., and R.W. wrote the manuscript. L.Z., V.S., J.Y., L.W., and R.W. designed the research. L.Z., I.M., S.D., and J.R. performed the research. L.Z., V.S., and K.K. analyzed the data. All authors have given final approval of the manuscript for submission.

1. Weinshilboum, R.M. & Wang, L. Pharmacogenomics: precision medicine and drug response. Mayo Clin. Proc. 92, 1711–1722 (2017).
2. Aldrich, S.L., Poweleit, E.A., Prows, C.A., Martin, L.J., Strawn, J.R. & Ramsey, L.B. Influence of CYP2C19 metabolizer status on esclotapram/citalopram tolerability and response in youth with anxiety and depressive disorders. Front. Pharmacol. 10, 99 (2019).
3. Hicks, J.K. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 genotypes and dosing of selective serotonin reuptake inhibitors. Clin. Pharmacol. Ther. 98, 127–134 (2015).
4. Sim, S.C. et al. A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. Clin. Pharmacol. Ther. 79, 103–113 (2006).
5. Veldic, M. et al. Cytochrome P450 2C19 poor metabolizer phenotype in treatment resistant depression: treatment and diagnostic implications. Front. Pharmacol. 10, 83 (2019).
6. Bielinski, S.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. 89, 25–33 (2014).
7. Chiasson, M., Dunham, M.J., Rettie, A.E. & Fowler, D.M. Applying multiplex assays to understand variation in pharmacogenes. Clin. Pharmacol. Ther. 106, 289–294 (2019).
8. Daly, A.K., Rettie, A.E., Fowler, D.M. & Miners, J.O. Pharmacogenomics of CYP2C9: functional and clinical considerations. J. Pers. Med. 8, pii: E17 (2017).
9. Devarajan, S. et al. Pharmacogenomic next-generation dna sequencing: lessons from the identification and functional characterization of variants of unknown significance in CYP2C9 and CYP2C19. Drug Metab. Dispos. 47, 425–435 (2019).
10. Vaaser, R., Adusumalli, S., Leng, S.N., Skic, M. & Ng, P.C. SiFT missense predictions for genomes. Nat. Protoc. 11, 1–9 (2016).
11. Adzhubei, I.A. et al. A method and server for predicting damaging missense mutations. Nat. Meth. 7, 248–249 (2010).
12. Choi, Y. & Chen, A.P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745–2747 (2015).
13. Kinney, J.B. & McCandlish, D.M. Massively parallel assays and quantitative sequence-function relationships. Annu. Rev. Genomics Hum. Genet. 20, 99–127 (2019).
14. Fowler, D.M. & Fields, S. Deep mutational scanning: a new style of protein science. Nat. Meth. 11, 801–807 (2014).
15. Matreyek, K.A., Stepanhan, J.J. & Fowler, D.M. A platform for functional assessment of large variant libraries in mammalian cells. Nucleic Acids Res. 45, e102 (2017).
16. Matreyek, K.A. et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat. Genet. 50, 874–882 (2018).
17. Wrenbeck, E.E., Klaasm, J.R., Stapleton, J.A., Adeniran, A., Tyn, K.E. & Whitehead, T.A. Plasmid-based one-pot saturation mutagenesis. Nat. Meth. 13, 926–930 (2016).
18. Leik, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016).
19. Stepanenko, A.A. & Dmitrenko, V.V. HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicy, and stress-induced genome-phenotype evolution. Gene 569, 182–190 (2015).
20. Gaedigk, A. et al. The Pharmacogene variation (PharmVar) consortium: incorporation of the human cytochrome P450 (CYP) allele nomenclature database. Clin. Pharmacol. Ther. 103, 399–401 (2018).
21. Wang, L., Nguyen, T.V., McLaughlin, R.W., Sikkink, L.A., Ramirez-Avila, M. & Weinshilboum, R.M. Human thiopurine S-methyltransferase pharmacogenetics: variant allele misfolding and aggresome formation. Proc. Natl. Acad. Sci. USA 102, 9394–9399 (2005).
22. Wang, L., Yee, V.C. & Weinshilboum, R.M. Aggresome formation and pharmacogenetics: sulfotransferase 1A3 as a model system. Biochem. Biophys. Res. Commun. 325, 426–433 (2004).
23. Li, F., Wang, L., Burgess, R.J. & Weinsilboum, R.M. Thiopurine S-methyltransferase pharmacogenetics: autophagy as a mechanism for variant allozyme degradation. Pharmacogenet. Genomics 16, 1083–1094 (2009).
24. Fowler, D.M., Stephany, J.J. & Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. Nat. Protoc. 9, 2267–2284 (2014).
25. Starita, L.M. et al. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. Proc. Natl. Acad. Sci. USA 110, E1263–E1272 (2013).
26. Gasperini, M., Starita, L. & Shendure, J. The power of multiplexed functional analysis of genetic variants. Nat. Protoc. 11, 1782–1787 (2016).
27. Wang, J.F., Wei, D.Q., Li, L., Zheng, S.Y., Li, Y.X. & Chou, K.C. 3D structure modeling of cytochrome P450 2C19 and its implication for personalized drug design. Biochem. Biophys. Res. Commun. 355, 513–519 (2007).
28. Reynald, R.L., Sansen, S., Stout, C.D. & Johnson, E.F. Structural characterization of human cytochrome P450 2C9: active site differences between P450s 2C8, 2C9, and 2C19. J. Biol. Chem. 287, 44581–44591 (2012).
29. Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J. Biol. Chem. 267, 83–90 (1992).
30. Rettie, A.E. & Jones, J.F. Clinical and toxicological relevance of CYP2C9: drug-drug interactions and pharmacogenetics. Annu. Rev. Pharmacol. Toxicol. 45, 477–494 (2005).
31. Flanagan, S.E., Patch, A.M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet. Test Mol. Biomarkers 14, 533–537 (2010).
32. Zi, J. et al. Effects of CYP2C9*3 and CYP2C9*13 on diclofenac metabolism and inhibition-based drug-drug interactions. Drug Metab. Pharmacokinet. 25, 343–350 (2010).
33. Chen, Y., Ferguson, S.S., Negishi, M. & Goldstein, J.A. Induction of human CYP2C9 by rifampicin, hyperforin, and phenobarbital is mediated by the pregnane X receptor. J. Pharmacol. Exp. Ther. 308, 495–501 (2004).
34. Sahi, J., Shord, S.S., Lindley, C., Ferguson, S. & LeCluyse, E.L. Regulation of cytochrome P450 2C9 expression in primary cultures of human hepatocytes. J. Biochem. Mol. Toxicol. 23, 43–58 (2009).
35. Dvorak, Z. & Paweł, P. Regulation of drug-metabolizing cytochrome P450 enzymes by glucocorticoids. Drug Metab. Rev. 42, 621–635 (2010).
36. Qin, S. et al. TSPYL family regulates CYP17A1 and CYP3A4 expression: potential mechanism contributing to abiraterone response in metastatic castration-resistant prostate cancer. Clin. Pharmacol. Ther. 104, 201–210 (2018).
37. Scott, S.A. et al. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450–2C19 (CYP2C19) genotype and clopidogrel therapy. Clin. Pharmacol. Ther. 90, 328–332 (2011).
38. Li-Wan-Po, A., Girard, T., Farrdon, P., Cooley, C. & Lithgow, J. Pharmacogenetics of CYP2C19: functional and clinical implications of a new variant CYP2C19*17. Br. J. Clin. Pharmacol. 69, 222–230 (2010).
39. Shirasaka, Y. et al. Interindividual variability of CYP2C19-catalyzed drug metabolism due to differences in gene diplotype and cytochrome P450 oxidoreductase content. Pharmacogenomics J. 16, 375–387 (2016).

© 2020 Mayo Clinic. Clinical and Translational Science published by Wiley Periodicals, Inc. on behalf of the American Society for Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.