Targeted Next-Generation Sequencing for Comprehensive Genetic Profiling of Pharmacogenes

SM Han[1], J Park[1], JH Lee[2], SS Lee[3], H Kim[4], H Han[4], Y Kim[4], S Yi[5], J-Y Cho[5], I-J Jang[5] and MG Lee[1]

Phenotypic differences in drug responses have been associated with known pharmacogenomic loci, but many remain to be characterized. Therefore, we developed next-generation sequencing (NGS) panels to enable broad and unbiased inspection of genes that are involved in pharmacokinetics (PKs) and pharmacodynamics (PDs). These panels feature repetitively optimized probes to capture up to 114 PK/PD-related genes with high coverage (99.6%) and accuracy (99.9%). Sequencing of a Korean cohort ($n = 376$) with the panels enabled profiling of actionable variants as well as rare variants of unknown functional consequences. Notably, variants that occurred at low frequency were enriched with likely protein-damaging variants and previously unreported variants. Furthermore, *in vitro* evaluation of four pharmacogenes, including cytochrome P450 2C19 (*CYP2C19*), confirmed that many of these rare variants have considerable functional impact. The present study suggests that targeted NGS panels are readily applicable platforms to facilitate comprehensive profiling of pharmacogenes, including common but also rare variants that warrant screening for personalized medicine.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ The genetic basis of phenotypic variations in drug responses has not been completely elucidated by existing PGx approaches with the limited number of available genetic markers.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ We demonstrated the performance of ADME-PGx panels as readily applicable targeted capture platforms. By applying these panels to a large cohort, genetic variants of PK/PD-associated genes were presented with regard to frequency spectra and functional impact.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE

☑ Despite recent large-scale sequencing projects, there are still a considerable number of novel variants, mostly occurring at low frequency. As predicted *in silico*, rare variants disturbed functions of proteins associated with ADME as shown for CYP2C19 and three transporters. In particular, novel or rare variants of *CYP2C19* exemplified under-rated variants of functional significance in drug phenotypes.

HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE

☑ Results suggest the importance of rare variants and the necessity of comprehensive PGx profiling. Targeted NGS-based panels will meet such needs both in preclinical research and clinical translation of pharmacogenomic discoveries.

Genetic variation among pharmacogenes greatly contributes to interindividual variability in drug response. Past pharmacogenomic studies have reached fruition, with implementation of well-characterized genetic variants for drug therapy decisions in clinics.[1–3] However, early pharmacogenomic platforms were confined to the analysis of common variants.[4,5] Consequently, there may be unreported, low-frequency variants, the effects of which are significant enough to be considered in determining therapeutic regimens. Indeed, approaches that focus exclusively on common genetic variants might have impeded clinical implementation of genetically tailored medicine by inadvertently excluding a potentially important class of variants and reducing their explanatory powers. There have been few studies showing rare variants with considerable functional impact. Variants with potential damaging consequences seem to be rare in populations[6] and play a role in drug-response phenotypes.[7] Even for drugs with an established genetic evidence-based dosing regimen, rare variants can increase the power to estimate the optimal dosing.[8]

Nevertheless, it is intrinsically more difficult and costlier to study the associations of rare variants than common variants due to the need for a large sample size. Therefore, the role of rare variants in variable drug-response phenotypes has not been fully evaluated,

[1]Department of Pharmacology, Yonsei University College of Medicine, Seoul, Korea; [2]Department of Clinical Pharmacology and Therapeutics, College of Medicine, Kyung Hee University, Seoul, Korea; [3]Department of Pharmacology and Pharmacogenomics Research Center, Inje University College of Medicine, Busan, Korea; [4]Celemics Inc., Seoul, Korea; [5]Department of Clinical Pharmacology and Therapeutics, Seoul National University College of Medicine and Hospital, Seoul, Korea. Correspondence: MG Lee (mlee@yuhs.ac)

Received 7 July 2016; accepted 27 September 2016; advance online publication 11 October 2016. doi:10.1002/cpt.532

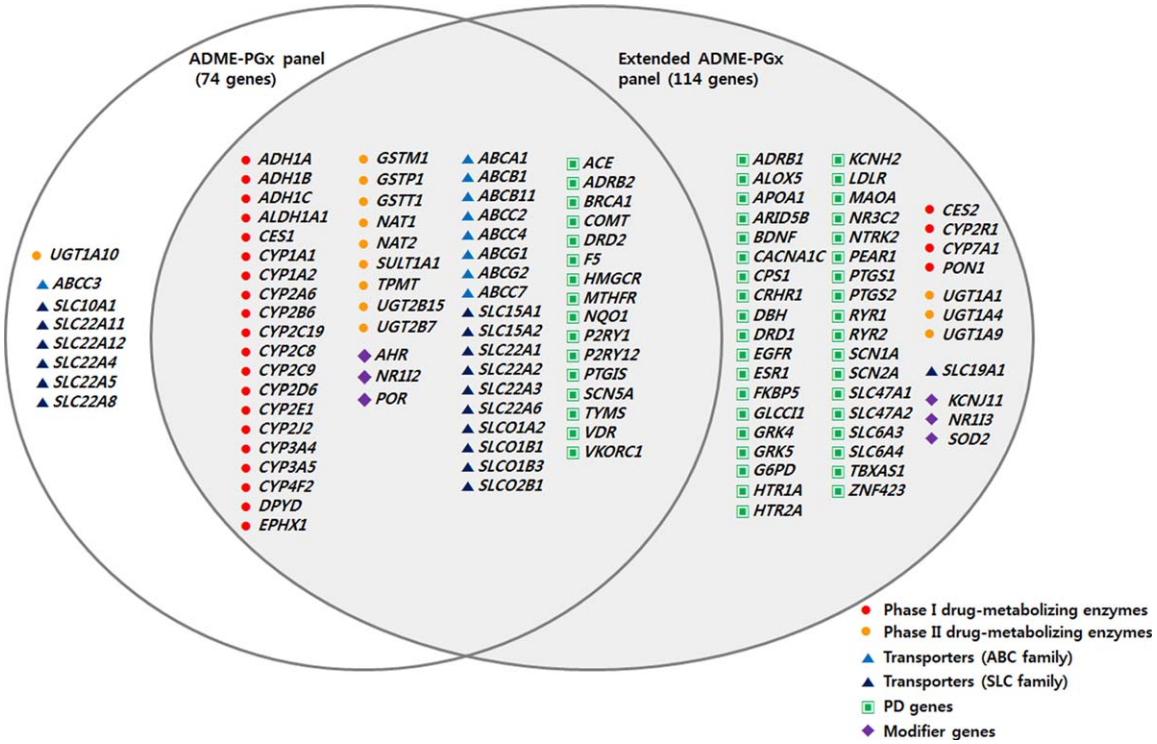

Figure 1 Target genes of the absorption, distribution, metabolism, and excretion (ADME)-pharmacogenetic (PGx) capture panels. Both ADME-PGx (n = 74 genes) and Extended ADME-PGx (n = 114 genes) panels capture pharmacokinetic (PK) genes of drug-metabolizing enzymes, transporters, modifiers, and pharmacodynamic (PD) genes. The Extended ADME-PGx panel includes more PD genes, which have been reported to be associated with drug response. Genes were categorized according to the PharmaADME database (http://www.pharmaadme.org/). ABC, ATP-binding cassette; SLC, solute carrier.

and their contributions in pharmacogenomics remain unclear. To examine the role of rare variants, there is a need for (1) efficient, comprehensive platforms that allow unbiased inspection of genetic variability in pharmacogenes; and (2) high-throughput molecular or *in silico* tools that can functionally validate variants.

Target-enrichment approaches have facilitated the efficient sequencing of genes of interest. Among three main target-capture approaches,[9] the molecular inversion probe (MIP)-based method has evolved into various versions with many improvements and been applied to large-scale sequencing.[10–12] The MIP-based approach enables capture with high specificity from small amounts of input DNA. Further, various modifications on the backbone sequence are possible to generate more efficient protocols, such as molecular tagging.[11,13] We previously suggested that the duplex MIP-based capture could be made less costly and laborious by synthesizing probes on a microarray.[13] By using this approach with a target-capture panel for 74 pharmacogenes (absorption, distribution, metabolism, and excretion (ADME)-pharmacogenetic (PGx)), we were able to circumvent the high cost of column-based probe synthesis and inaccurate genotype calling due to allelic bias. For an additional target-capture panel (extended ADME-PGx), we used in-solution hybridization capture to scale up the number of target genes. Hybridization capture by RNA probes in solution has been increasingly used for commercial exome capture kits.[14,15]

Here, by applying current target-enrichment strategies, we generated next-generation sequencing (NGS)-based target capture panels to sequence genes involved in pharmacokinetics (PKs) and pharmacodynamics (PDs). Probes consisting of these panels were sufficiently optimized to capture and sequence almost all target bases with adequate sequencing depth. By applying these panels to a large Korean cohort (n = 376), genetic variants in pharmacogenes were elucidated with regard to expected functional consequences and frequency in the population. In addition, the functional consequences of rare and previously unreported variants in the drug-metabolizing enzyme cytochrome P450 2C19 (*CYP2C19*) gene and three transporter genes (*SLC22A1*, *SLCO1B1*, and *ABCB1*) were validated *in vitro*. Collectively, our data suggest that these NGS panels are readily applicable platforms that enable comprehensive profiling of pharmacogenes. These NGS-based approaches will provide fertile ground for future applications of pharmacogenomics in precision medicine and drug development by considering rare variants that may contribute to interindividual variations in drug response.

RESULTS

NGS-based target capture panels for pharmacogenes

Two kinds of NGS-based target capture panels were generated to sequence PK/PD-related genes. Both panels covered 600 bp upstream and 100 bp downstream of untranslated regions and -50 bp/+10 bp of each splice junction of the target genes. Probes targeting markers of Affymetrix' DMET Plus array were also included in panels. Target genes of two ADME-PGx panels are presented in **Figure 1**. Initially, we developed an MIP capture-based ADME-PGx panel for 74 genes. The panel focused mainly on genes involved in ADME processes, including genes encoding

Table 1 Overall performance of two ADME-PGx panels

	ADME-PGx panel ($n = 74$ genes)	Extended ADME-PGx panel ($n = 114$ genes)
Captured samples, no.	191	185
Mean depth	400×	455×
Plex-level	20	40
Covered fraction ≥1×, %, mean	96.1	99.8
Covered fraction ≥15×, %, mean	91.5	99.6
Covered fraction ≥20×, %, mean	90.5	99.6
Precision, %[a]	99.8	99.9
Sensitivity, %[a]	98.1	99.7
Filter condition of genotype calls	Coverage (×) ≥15×, VAF ≥0.3	VAF ≥0.29
Concordance rate with DMET Plus assay, %[b]	99.4 (mean)	NA

ADME-PGx, absorption, distribution, metabolism, and excretion-pharmacogenetic; NA; not available; Plex-level, number of samples pooled per lane in the Hiseq 2500 sequencing system; Precision, true positive variant calls by panels among all variants from next generation sequencing (NGS); Sensitivity, true positive variant calls within covered regions by panels among the benchmark variant call set; VAF, variant allelic fraction, meaning the fraction of sequence reads carrying variant alleles.

[a]Precision and sensitivity of genotype calls by two panels were measured by comparing genotypes of NA12878 to the benchmark genotype calls. Although a total of 486 single nucleotide variations (SNVs) and in-del genotypes were available for comparison for ADME-PGX, a total of 685 SNVs and in-del genotypes were compared for Extended ADME-PGX panel. [b]The concordance rate was derived from the data of 191 samples for 1,831 markers that passed quality controls for the DMET Plus assay (call rate ≥98%) and the post-filter for NGS variant calls (sequencing depth ≥15× and VAF ≥0.3) by the ADME-PGx panel.

phase I/II drug-metabolizing enzymes, drug transporters, and modifiers of other ADME genes. Compared with existing methods, the MIP-based capture used in the panel featured efficient probe preparation by adopting microarray-based duplex MIP with minor modifications.[13] We initially designed 10,927 probes and repetitively optimized them in order to improve capture performance (**Supplementary Figure S1a** online). Melting temperature of the annealing sequence for ligation and extension is known to be one predictor of capture efficiency and can be easily adjusted.[10] The ideal range of melting temperatures of two annealing sequences was empirically defined at the pilot stage and ranged from 55–62.5°C. We found that even one probe with adequate target capture performance (ligation or extension) could overcome the poor capture efficiency of the other probe (**Supplementary Figure S1b** online). After the initial run, a total of 2,099 MIPs, target regions of which were covered by less than the mean 25× sequencing coverage or for which melting temperatures of both annealing arms were out of the optimal range, were specified as probes with poor capture performance (**Supplementary Figure S1c** online). To complement those, we generated 3,206 new probes whose melting temperatures fell within the optimal range. The newly designed probes showed increased coverage over target regions (**Supplementary Figure S1d** online).

In addition to the MIP-based platform, we generated the Extended ADME-PGx panel covering 114 pharmacogenes based on hybridization capture using RNA probes in order to compare the capture performance of the two platforms and to establish the best NGS platform for PGx testing. We optimized the probes by repetitively measuring sequencing coverage and the uniformity of each probe. Regions that were less uniform due to repeat sequences or high guanine-cytosine (GC) or adenine–thymine (AT) ratios were chosen to be rebalanced. These rebalancing procedures improved the capture performance of the Extended ADME-PGx panel. For example, coverage uniformity at 0.2× of mean depth of coverage was improved from 92.3% to 97.3% after rebalancing (**Supplementary Figure S2** online).

Overall performance of capture panels

Optimized MIPs (ADME-PGx) captured 96.1% of target genome sequences and 91.5% at sequencing depth of 15×, which met the minimum requirement of variant calls for high precision and sensitivity (**Table 1**, **Supplementary Figure S3a,b** online). Covered regions were sequenced to the mean depth of 400×. At minimum conditions of accurate and sensitive variant calling, genotyping of NA12878 by the ADME-PGx panel was 99.8% concordant with validated genotypes with 98.1% sensitivity (**Table 1**). When compared to genotyping using the commercial genotype platform, the DMET Plus panel, genotypes called by the ADME-PGx panel showed 99.4% concordance (**Table 1**).

The hybridization-based Extended ADME-PGx panel showed better overall capture performance. Greater than 99% of target bases were covered at sequencing depth of 20× by the extended panel. By sequencing target regions with the mean depth of 455×, the Extended ADME-PGx panel captured nearly all target bases of 114 genes (**Figure 2**). With regard to variant calls from the hybrid-capture, filtering based on variant allele fraction (VAF), but not filtering based on sequencing depth (×), improves precision and sensitivity (**Supplementary Figure S3c,d** online). The Extended ADME-PGx panel provided 99.9% accurate variant calls of NA12878 with 99.7% sensitivity (**Table 1**). When the sequencing depth of both panels were normalized at 375×, again the hybridization-based Extended ADME-PGx panel showed better performance than the MIP-based panel in the present dataset (**Supplementary Figure S4** online).

Analysis of variants in pharmacogenes

We sequenced the DNA of 376 healthy Koreans with two ADME-PGx panels. Among 376 individuals, sequencing of 191 individuals was performed with the MIP-based ADME-PGx panel. For the remaining 185 individuals, the Extended ADME-PGx panel was applied. There was no sample overlap for target captures by the two panels. The number of variants per gene, their minor allele frequency (MAF), and novelty are presented in **Figure 3** (Extended ADME-PGx) and **Supplementary Figure S5** online (ADME-PGx). The overall statistics of the identified variants are provided in **Supplementary Table S1** online. Of the genes sequenced, *RYR1* had the largest number of

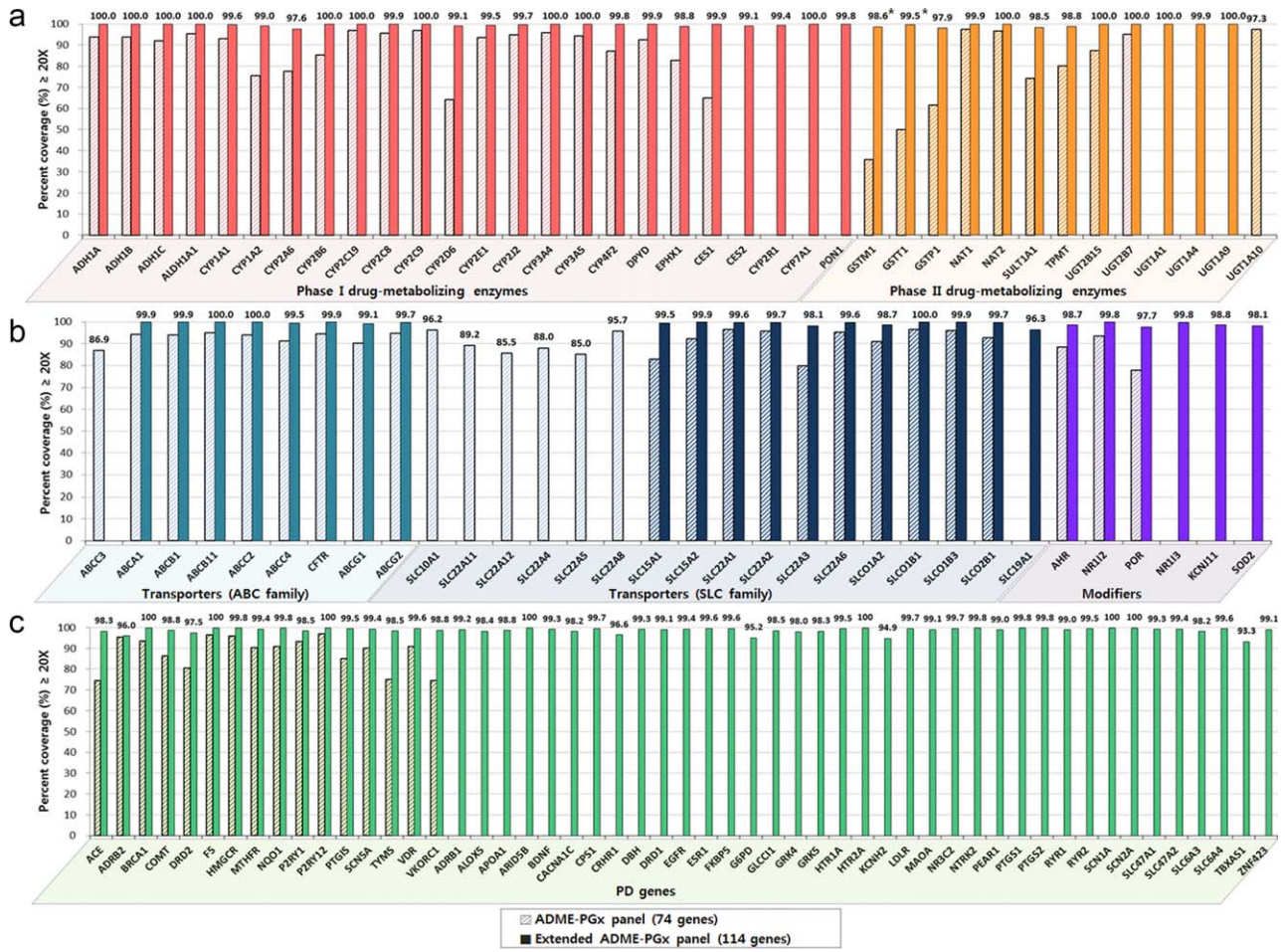

Figure 2 Percentage of bases with a minimum of 20× coverage by target capture panels. Percentage of bases with >20× coverage by the ADME-PGx (hatched) and Extended ADME-PGx (filled) panels is shown for six classes of target genes (**a–c**). Using the Extended ADME-PGx panel, >99% of target bases for 114 genes were captured and sequenced. Bars representing the greater coverage of the two panels are labeled with the numeric value. *Mean coverage (%) was calculated without carriers of nonfunctional alleles (GSTT1*0 and GSTM1*0) as confirmed by the polymerase chain reaction-based method (see **Supplementary Figure S9** online). ABC, ATP-binding cassette; SLC, solute carrier.

variants. However, when the number of variations was normalized to its gene length, *RYR1* was not a highly variable gene for its size (**Supplementary Figure S6** online). Similarly, ATP-binding cassette (ABC) family transporters were likely to have more variants compared to solute carrier (SLC) family, (**Figure 3** and **Supplementary Figure S5** online), due to their large gene size (**Supplementary Figure S6** online), as shown by a recent study.[16] Interestingly, *CYP2R1* harbored the least number of variants among cytochrome P450 (CYPs), as reported earlier.[17] Considering that CYP2R1 has an essential role in vitamin D metabolism,[18] it seems that variants of *CYP2R1* have been under natural selections.

One parameter that might provide insights into the sequencing results of pharmacogenes is the DMET Plus coverage (variants covered by the DMET Plus array/total variants identified in the NGS panel for each DNA sample, %). Of the 231 genes on the DMET Plus array, 61 genes were targeted by the ADME-PGx panel, and 64 genes were targeted by the Extended ADME-PGx panel. Average DMET Plus coverage of ADME-PGx and Extended ADME-PGx was 33.9% and 31.1%, respectively, which is consistent with a recent study reporting that DMET Plus coverage of a whole-genome sequencing-based analysis was ~26.4–40.0%, depending on the samples.[19]

After normalizing the number of variants by their gene size, *CYP2A6* and *CYP2D6* had the highest rate of genetic variability (**Supplementary Figure S6** online), which is in line with previous studies.[17,20] However, due to the high complexity of these loci, as listed in problematic regions for short read NGS approach,[21] we cannot completely exclude false-positive variant calls derived from misalignment of short reads. Therefore, we further examined our sequencing results on *CYP2D6* using the following methods. First, the sequencing results of 191 individuals in the ADME-PGx panel were compared with 30 target loci of *CYP2D6* in the DMET Plus panel. The targeted sequencing of *CYP2D6* by ADME-PGx showed 99.7% concordance with the DMET Plus panel, suggesting that results between our genotyping and DMET plus panels are reproducible on *CYP2D6* loci. Second, genotyping data of the reference sample NA12878 by Extended ADME-PGx were compared with those from open databases composed of five NGS-based platforms and a Sanger-

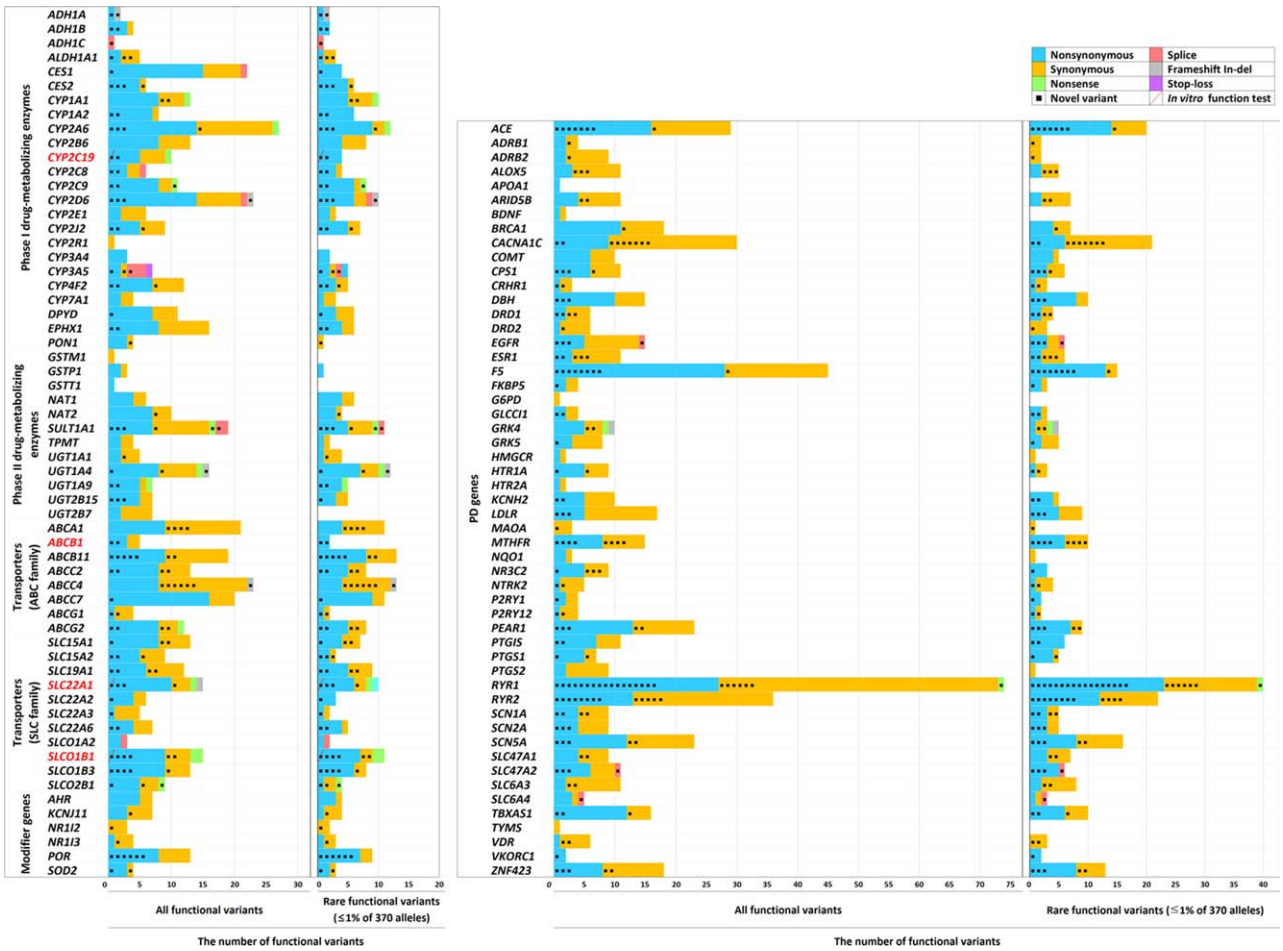

Figure 3 The overview of the number of functional variants in 114 pharmacogenes. Summarized data of DNA samples from 185 individuals, captured by the Extended ADME-PGx panel, are shown. Functional variants were defined as nonsynonymous (shown in light blue), synonymous (shown in orange), nonsense (shown in green), splice-site (shown in red), stop-loss (shown in purple) variants, and frameshift insertion and deletion (shown in gray). Novel variants were defined based on database-single-nucleotide polymorphism (dbSNP) build 142 (black square) and variants that were evaluated *in vitro* functional tests were shown with red slash. Rare variants (minor allele frequency (MAF) ≤1%, less than three alleles among 370 alleles) are separately shown in each right panel. ABC, ATP-binding cassette; SLC, solute carrier.

based *CYP2D6* genotyping dataset[22] (**Supplementary Table S2** online). Of the 14 loci analyzed, genotyping on a *CYP2D6* locus (rs1058172) did not agree with that of the Sanger validation. Although our Extended ADME-PGx panel showed the highest concordance rate with the Sanger genotyping data (13/14; 92.9%) among the currently available NGS-based data sets (**Supplementary Table S2** online), this result indicates that there is still room for improvement in NGS-based genotyping of high-complexity regions.

Novel and rare functional variants across healthy Koreans

We identified novel variants of PK/PD-associated genes that have not been reported in large sequencing projects, such as the 1000 Genome Project and the Exome Aggregation Consortium. When predicting functional effects of variants by *in silico*, functional prediction programs, such as SIFT, PolyPhen-2, and Combined Annotation Dependent Depletion (CADD) rare variants showed a tendency to have detrimental effects on protein functions (**Figure 4**). Most of these variants occurred at low frequency and showed varying degrees of damaging effects predicted by the three prediction algorithms. To provide insight into novel and rare variants, the functional consequences of variations in four clinically important PK genes (*CYP1C19*, *SLC22A1*, *SLCO1B1*, and *ABCB1*; **Supplementary Figure S7** online) were examined *in vitro*. Variants were selected using the following criteria: (1) nonsynonymous variant with MAF <0.5%; (2) functionality not previously reported in the literature or databases; and (3) predicted to have damaging or deleterious effects by more than two *in silico* prediction algorithms. We initially examined variants of *CYP2C19*, which is one of the most frequently listed genes in the US Food and Drug Administration table of pharmacogenomic biomarkers.[23] We have identified seven nonsynonymous *CYP2C19* variants across 376 individuals (**Table 2**, **Supplementary Table S3** online). Not only common actionable variants, such as p.W212X (CYP2C19*3), but also unreported and novel variants of unknown function were found. To verify actual functional changes by novel rare genetic variants, we investigated *in vitro* functional changes of two variants in *CYP2C19*

Table 2 Nonsynonymous variants of *CYP2C19* including common and rare variants ($n = 376$)

Genomic coordinate	Nucleotide change (transcript version)	Amino acid change	Minor allele frequency ($n = 376$)	1000 GP	ExAC	No. of algorithms predicting as damaging[a]	*In vitro* function
chr10:96541753	c.818A>C (NM_000769)	p.K273T	0.001	Novel	Novel	3/6	Significant decrease (in this study)
chr10:96522541	c.79G>C (NM_000769)	p.G27R	0.001	Novel	Novel	1/6	Not investigated
chr10:96540403	c.629C>A (NM_000769)	p.T210N	0.003	Novel	AF = 0.00002 (3/121142)	3/6	Significant decrease (in this study)
chr10:96580264	c.831C>A (NM_000769)	p.N277K	0.003	rs559628884	AF = 0.00098 (12/121362)	1/6	No significant difference[44]
chr10:96535296	c.481G>C (NM_000769)	p.A161P	0.004	rs181297724	AF = 0.0048 (581/121408)	4/6	Significant decrease[45]
chr10:96602623	c.991G>A (NM_000769)	p.V331I	0.048	rs3758581	AF = 0.0624 (7577/121394)	0/6	No significant difference[46]
chr10:96540410	c.636G>A (NM_000769)	p.W212X	0.102	rs4986893	AF = 0.0055 (667/121120)	2/6	Significant decrease

1000 GP, 1000 Genomes Project; ExAC, Exome Aggregation Consortium.
Target variants for *in vitro* investigation were confined to only rare (minor allele frequency <0.005) or novel variants with predicted damaging impacts but unknown functional consequence: these were p.K273T and p.T210N.
[a]PolyPhen-2, SIFT, MutationTaster, Scaled C-scores, MutationAssessor, and I-Mutant 2.0 algorithms were applied for *in silico* functional prediction.

(p.T210N and p.K273T). These two variants were predicted to be deleterious by most *in silico* prediction algorithms but were never investigated regarding actual impacts on the protein. Both variants significantly lowered CYP2C19 activity, which is responsible for 4′-hydroxylation of (S)-mephenytoin (**Figure 5**). In addition, we investigated *in vitro* functional impacts of rare or novel variants of three transporters (SLC22A1/OCT1, SLCO1B1/OATP1B1, and ABCB1/MDR1) on reference substrates (**Supplementary Table S4** online). The three transporters have been reported to be clinically relevant to several drugs with varying strength of evidence (**Supplementary Figure S7** online). Most functional target variants of the three transporter genes significantly impaired the intrinsic functions of the proteins. Three SLC22A1/OCT1 variants (p.A33V, p.Y91C, and p.A407T) significantly decreased uptake of (^{3}H)-methyl-4-phenylpyridinium and (^{14}C)-tetraethylammonium (**Supplementary Figure S8** online). These results were consistent with predictions by most computational algorithms (**Supplementary Table S4** online). The functional consequences of two SLCO1B1/OATP1B1 variants (p.V84L and p.V405F) were also in line with most *in silico* predictions. The p.V84L variant significantly decreased the cellular influx of (^{3}H)-estrone-3-sulfate (**Supplementary Table S4, Supplementary Figure S8** online). Among three likely protein-damaging variants of ABCB1/MDR1, two variants (p.G251R and p.L760V) decreased the MDR1-mediated calcein efflux in HEK 293T cells

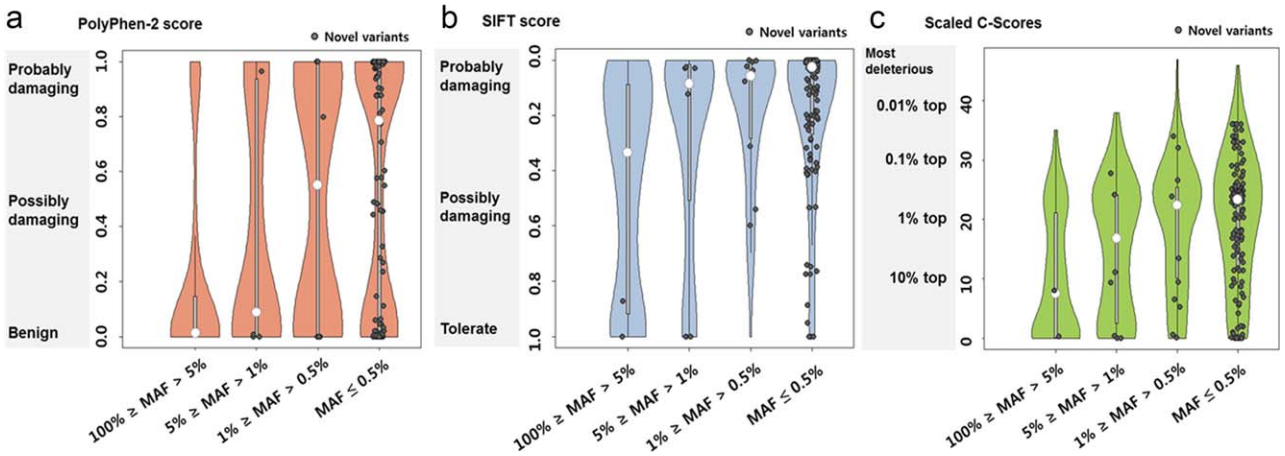

Figure 4 Functional prediction of genetic variants shown by minor allele frequency (MAF) and novelty in 376 healthy Koreans. Expected functional consequences of genetic variants in target pharmacokinetic/pharmacodynamic (PK/PD) genes were plotted with regard to MAF in the population and novelty. Predictions were based on three different *in silico* prediction algorithms, PolyPhen-2 (**a**), SIFT (**b**), and Combined Annotation Dependent Depletion (CADD) (**c**). Variants that could not be analyzed by PolyPhen-2 and SIFT, such as nonsense variants, were not included in the plots. Results were based on data from genes targeted by ADME-PGx (191 individuals) and Extended ADME-PGx panels (185 individuals).

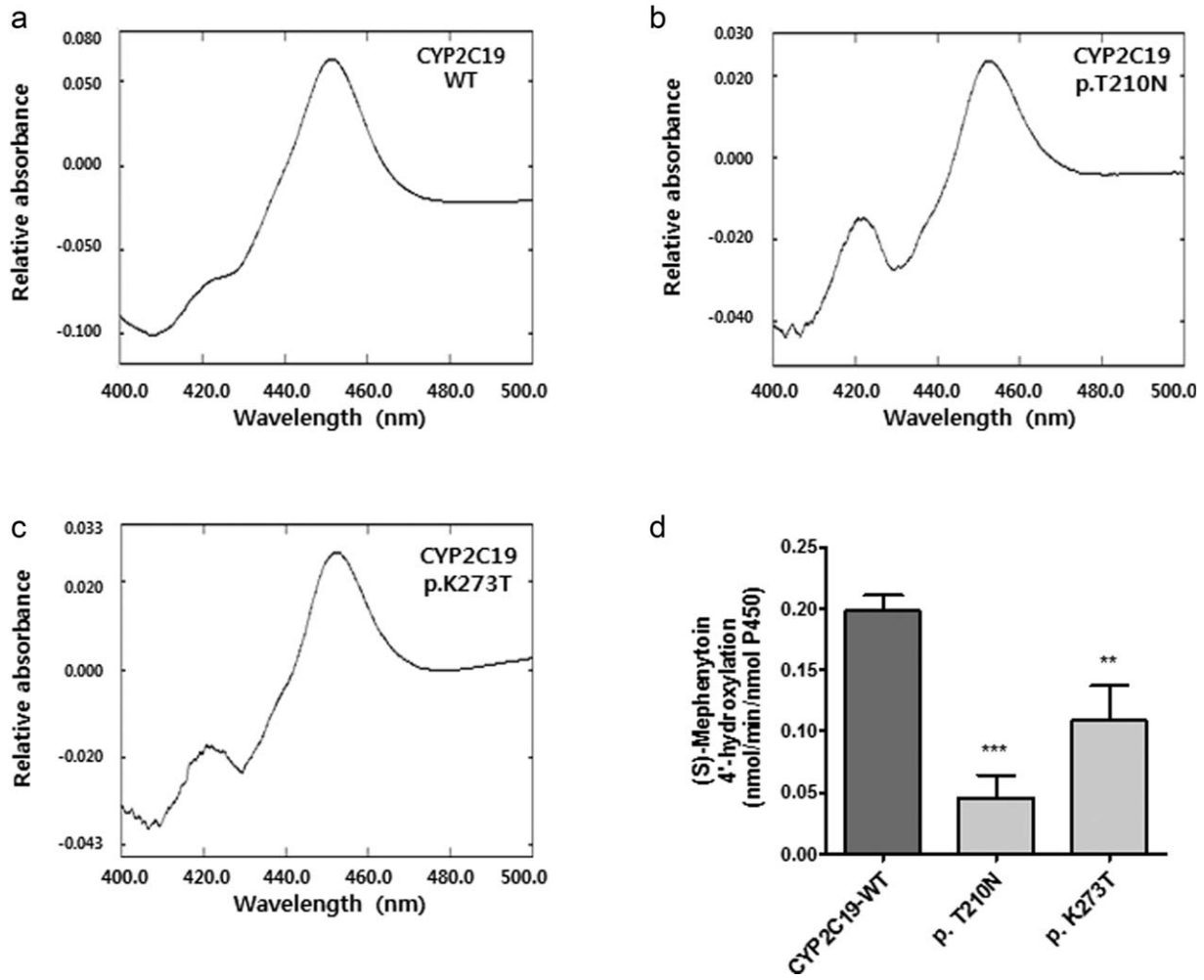

Figure 5 Reduced carbon monoxide (CO)-difference spectra and functional consequences of rare variants of cytochrome P450 2C19 (CYP2C19). (**a–c**) Reduced CO-difference spectra in an *E. coli* system were monitored for wild-type (WT) and mutant CYP2C19s (p.T210N and p.K273T). The P450 content was determined by measuring the maximum CO-reduced spectrum at 450 nm. Representative spectra for measuring P450 content of protein samples of WT (**a**), p.T210N (**b**), and p.K273T (**c**) CYP2C19s are shown. (**d**) Relative enzymatic activity of WT and two mutated forms of CYP2C19 for 4′-hydroxylation of (S)-mephenytoin ($n = 6$). Results are expressed as mean ± SEM. $^{**}P < 0.01$, $^{***}P < 0.001$ relative to WT activity.

(**Supplementary Table S4, Supplementary Figure S8** online). These data show that, although these novel variants in core PK/PD-related genes are present at low frequency, they have considerable functional impact in pharmacotherapy.

DISCUSSION

We developed NGS-based capture panels that consist of optimized probe sets based on two different capture techniques. The overall performance of the panels was acceptable for genetic studies in terms of precision and sensitivity (**Table 1**). As the panels were developed with flexible addition and deletion of probes, these platforms will be useful for various PGx testing applications after some modifications. Notably, the hybridization-based Extended ADME-PGx panel seems to be a readily applicable platform that meets the need of high performance for clinical sequencing. Although MIP-based methods are ideally more specific to capture targets, this feature would make it more difficult to capture GC-rich or AT-rich regions or repeated sequences. In contrast, even with less stringent filtering thresholds, the hybrid-capture data were sufficient for precise variant calls once they reached a significant depth (>20×). We expect that, especially for diagnostic purposes, use of the hybrid-capture for targeted NGS would continue for a while as the most efficient alternative to whole-genome sequencing.

Meanwhile, genes encoding glutathione transferase enzymes, such as GSTM1 and GSTT1, were not evenly captured by both panels despite repetitive rebalancing. We speculated that common null functional alleles of both genes resulted in the failure to capture such regions and further confirmed this by a polymerase chain reaction-based method[24] (**Supplementary Figure S9** online). We identified that null alleles of two genes (GSTT1*0 and GSTM1*0) are quite common in the Korean population, in which allele frequencies were 0.71 and 0.77, respectively. When counting only samples possessing more than one copy of *GSTM1* and *GSTT1*, almost all bases were sufficiently sequenced through the extended panel. Interestingly, the number of copies could be predicted from the mean number of sequencing reads, which was

proportional to copy number confirmed by the polymerase chain reaction-based method (**Supplementary Figure S9** online).

The NGS-based investigation of relevant pharmacogenes and drug-response phenotypes is still in its beginning stages. Recent efforts have identified novel and actionable gene variants from thousands of patients through sequencing-based panels.[25] The results show that a considerable proportion of patients harbor potentially damaging rare gene variants relevant to frequently prescribed drugs. By using capture panels to analyze the DNA of 376 Koreans, we identified a number of unreported genetic variants, including those occurring at low frequency. Pathogenicity prediction programs suggest that rare genetic variants are likely to have deleterious effects on protein function (**Figure 4**). We compared the performance of six *in silico* algorithms with results of an *in vitro* function test (**Supplementary Table S4** online). Interestingly, the predictions of algorithms based on evolutionary conservation, such as SIFT,[26] MutationTaster,[27] and CADD scaled C scores,[28] seemed to be more accurate than those of algorithms incorporating structural stability, such as PolyPhen-2,[29] Mutation Assessor,[30] and I-Mutant 2.0.[31] These results may be due to insufficient information on the relevant protein structure or local structural motif of unknown genetic variations. However, the limited number of gene variants examined *in vitro* in the present study were not sufficient to evaluate the overall performance of the six *in silico* algorithms. A more comprehensive comparison between *in silico* algorithms and *in vitro* function is required to draw a valid conclusion. However, the current versions of the better performing algorithms (SIFT, MutationTaster, and CADD) predict functional consequences of rare variations with an accuracy of ~80% (**Supplementary Table S4** online). This reinforces the need to develop a high-throughput molecular method or clinically acceptable algorithm to estimate the effects of rare genetic variations on PK/PD in individual patients. Of interest, when all three algorithms made consistent predictions, the results matched those of the *in vitro* assay.

The *in vitro* evaluation of four PK genes in this study indicated that low-frequency genetic variants can have considerable functional impact. For instance, we found that two rare *CYP2C19* variants (p.T210N and p.K273T) significantly decreased enzymatic activity (**Figure 5**, **Table 2**). Because these two sites are not within the specific binding sites of (S)-mephenytoin,[32] they seem to change the catalytic efficiency of the enzyme. In *OCT1*, three nonsynonymous rare variants (p.A33V, p.Y91C, and p.A407T) decreased transmembrane transport activity of the protein (**Supplementary Table S4** online), which may affect the PK of several clinically important drugs, such as metformin.[33] Reduced OATP1B1 activity by common genetic variants is associated with statin-induced myopathy,[34] suggesting that decreased OATP1B1 function by the rare variant p.V84L (**Supplementary Table S4** online) may also increase the risk of statin-induced adverse drug reactions. MDR1 is one of the most widely distributed proteins in the body and is involved in PK of many drug molecules.[35] Consequently, decreased MDR1 function by p.G251R and p.L760V may influence the absorption and elimination of MDR1 substrate drugs. Interestingly, a different MDR1 allele at Gly251 (p.G251V) was previously reported as a processing mutant with decreased activity.[36] Although both p.G251V and p.G251R variants decrease MDR1 function, only the known variant p.G251V is listed in conventional genotyping platforms, such as the DMET Plus panel. These results highlight the utility of NGS-based methods even for known pharmacogenomic loci.

Clinical implementation of NGS-based tests requires appropriate turnaround times and cost-effectiveness.[37] This will be one of the priority issues to integrate NGS-based genetic profiling for clinical therapeutics. Although rapid progress in sequencing technology will overcome practical challenges of whole exome and genome analyses, targeted NGS panels might be suitable for genetic testing at present to satisfy the requirements of clinically applicable turnaround times and cost-effectiveness. In addition, a number of pilot models integrating preemptive pharmacogenomic testing have validated its utility.[38–40] Multiplex genotyping in advance could reduce the overall number of tests and, therefore, was expected to be more cost-efficient.[41] Preemptive testing with NGS-based panels would enable the further investigation of variants with unknown functional consequences for novel pharmacogenomics discovery.[42] Furthermore, we expect that NGS-based genetic profiling at an earlier phase of clinical trials will provide the foundation for understanding unknown metabolic pathways of drugs by taking rare variants into account. In conclusion, novel and rare variants of *in vitro* functional significance in our cohort imply the necessity of comprehensive genetic screening. We suggest that targeted NGS panels are readily applicable platforms both in preclinical and clinical settings for unbiased profiling of pharmacogenes.

METHODS

Sample collection

Genomic DNA (gDNA) was obtained from peripheral whole blood samples of 376 healthy Korean male volunteers at Seoul National University Hospital. Whole blood samples were anonymized and stored at temperatures below -70°C until gDNA extraction. DNA was isolated using the Gentra Puregene DNA Isolation Kit (Qiagen, Hilden, Germany). All participants gave written informed consent prior to sample collection, which was approved by an Institutional Review Board of Seoul National University Hospital (project number: H-0803-022-237) and Yonsei University College of Medicine (project number: 1-2015-0056). All procedures were performed in compliance with guidelines for the Declaration of Helsinki. Of the 376 individuals, 191 individuals were genotyped first by using the DMET Plus panel and then with the MIP-based ADME-PGx panel. The DNA of the remaining 185 individuals was sequenced by using the hybrid-capture method with the Extended ADME-PGx panel. The gDNA from HapMap sample NA12878 was purchased from the Coriell Institute and used as a reference.

Design, generation, and optimization of probes

Probes were designed by fetching sequences of desired length from the reference genome sequences (hg19/Build 37), while maintaining overlap of ends of adjacent probes. For the MIP-based approach, we assigned annealing sequences within the ideal range of melting temperatures for each probe, determined at the pilot stage. To circumvent genotyping errors due to polymerase chain reaction bias of the MIP-based capture, we added 15 randomly distributed bases as molecular tags to be used to eliminate duplicated reads. Microchip-based synthesis of oligonucleotides was followed by amplification to obtain a sufficient amount of probes to capture target genes from hundreds of genomes. For even and efficient

MIP amplification, emulsion polymerase chain reaction was carried out following minor modification of the previous protocol.[43] From high-throughput sequencing data, we specified probes of poor capture efficiency by measuring the mean sequencing depth of 16 samples. MIPs, target regions of which were covered by <25×, or for which melting temperatures for both annealing arms were out of the optimal range, were redesigned. New probes were designed by adjusting the length of annealing sequences or shifting the location of annealing sequences from GC-rich or AT-rich regions. Probes with low capture performance but within the optimal range of melting temperatures were increased in amount. Redesigned probes were simply mixed with the original probes at a 1:1 ratio after verifying the presence of negligible competition between probes.

For hybridization capture in the Extended ADME-PGx panel, target regions of 114 genes were defined by merging all transcript isoforms and nearby regions. Two different sets of oligonucleotide probes were initially generated by shifting positions within target regions and their performance was tested using reference samples (NA12878, NA12891, and NA12892). After evaluating the target-capture efficiency of the initial probe sets, inefficient probes were increased in concentration or replaced with redesigned probes. These rebalanced probes showed the performance of 97.3% uniformity (at 0.2× mean depth of coverage) and 98.8% coverage (at 20× depth) (**Supplementary Figure S2** online). The final probe combination was determined through repetitive rebalancing procedures, and then the final probe sets were chemically synthesized as a single pool of optimized probes. These probes, which comprise the Extended ADME-PGx panel, exhibited 99.6% coverage (at 20× depth; **Table 1**).

Target-capture and high-throughput sequencing

Target-capture sequencing using the ADME-PGx panel was performed as described previously with slight modifications.[13] A total of 1 μg gDNA was used for the MIP-based capture (gDNA:probe ratio of 1:100). Sequencing was performed in a 20-plex assay for the MIP-based ADME-PGx panel. For target-capture sequencing in the Extended ADME-PGx panel, we sheared 0.2–1 μg gDNA (150–250 bp) using the Diagenode Bioruptor Pico Ultrasonicator. We performed 40-plex sequencing for the hybrid capture-based Extended ADME-PGx panel. Additional details of procedures for target-capture and NGS sequencing are described in **Supplementary Supporting Information** online.

Bioinformatics and in-house programs for genotype calling

See **Supplementary Supporting Information** online.

Evaluation of capture performance and genotyping of *GSTT1* and *GSTM1* null genotype

See **Supplementary Supporting Information** online.

Selection of variants for *in vitro* evaluation and measurements of CYP2C19, OCT1, OATP1B1, and MDR1 activities

See **Supplementary Supporting Information** online.

Statistical analysis

For results of *in vitro* assays, data are presented as mean ± SEM from at least three experiments. Statistical analysis was performed using Student's *t*-tests or analysis of variance, as appropriate. Any $P < 0.05$ was considered as statistically significant.

Additional Supporting Information may be found in the online version of this article.

ACKNOWLEDGMENTS

This research was supported by grants from the National Research Foundation of Korea (No. NRF-2013R1A3A2042197), and from the Korea Health Technology R&D Project, Korea Health Industry Development Institute (KHIDI), Ministry of Health & Welfare (No. HI15C1543), Republic of Korea. M.G.L. was supported by the faculty research programs (6-2008-0233, 6-2009-0152) of Yonsei University College of Medicine.

CONFLICT OF INTEREST

The authors declared no conflict of interest.

AUTHOR CONTRIBUTIONS

S.M.H. and M.G.L. wrote the manuscript. S.M.H., J.P., J.H.L., J.-Y.C., I.J., and M.G.L. designed the research. S.M.H., J.P., S.S.L., H.K., and H.H. performed the research. S.M.H., S.Y., and Y.K. analyzed the data.

1. Daly, A.K. Genome-wide association studies in pharmacogenomics. *Nat. Rev. Genet.* **11**, 241–246 (2010).
2. Motsinger-Reif, A.A. *et al*. Genome-wide association studies in pharmacogenomics: successes and lessons. *Pharmacogenet. Genomics* **23**, 383–394 (2013).
3. Ventola, C.L. Pharmacogenomics in clinical practice: reality and expectations. *P T* **36**, 412–450 (2011).
4. Burmester, J.K., Sedova, M., Shapero, M.H. & Mansfield, E. DMET microarray technology for pharmacogenomics-based personalized medicine. *Methods Mol. Biol.* **632**, 99–124 (2010).
5. Daly, T.M. *et al*. Multiplex assay for comprehensive genotyping of genes involved in drug metabolism, excretion, and transport. *Clin. Chem.* **53**, 1222–1230 (2007).
6. Nelson, M.R. *et al*. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100–104 (2012).
7. Ramsey, L.B. *et al*. Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. *Genome Res.* **22**, 1–8 (2012).
8. Sagreiya, H. *et al*. Extending and evaluating a warfarin dosing algorithm that includes CYP4F2 and pooled rare variants of CYP2C9. *Pharmacogenet. Genomics* **20**, 407–413 (2010).
9. Mamanova, L. *et al*. Target-enrichment strategies for next-generation sequencing. *Nat. Methods* **7**, 111–118 (2010).
10. O'Roak, B.J. *et al*. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
11. Hiatt, J.B., Pritchard, C.C., Salipante, S.J., O'Roak, B.J. & Shendure, J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res.* **23**, 843–854 (2013).
12. Shen, P., Wang, W., Chi, A.K., Fan, Y., Davis, R.W. & Scharfe, C. Multiplex target capture with double-stranded DNA probes. *Genome Med.* **5**, 50 (2013).
13. Yoon, J.K. *et al*. microDuMIP: target-enrichment technique for microarray-based duplex molecular inversion probes. *Nucleic Acids Res.* **43**, e28 (2015).
14. Parla, J.S., Iossifov, I., Grabill, I., Spector, M.S., Kramer, M. & McCombie, W.R. A comparative analysis of exome capture. *Genome Biol.* **12**, R97 (2011).
15. Chilamakuri, C.S. *et al*. Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics* **15**, 449 (2014).
16. Kozyra, M., Ingelman-Sundberg, M. & Lauschke, V.M. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet. Med.* (2016); e-pub ahead of print.
17. Fujikura, K., Ingelman-Sundberg, M. & Lauschke, V.M. Genetic variation in the human cytochrome P450 supergene family. *Pharmacogenet. Genomics* **25**, 584–594 (2015).
18. Zhu, J.G., Ochalek, J.T., Kaufmann, M., Jones, G. & Deluca, H.F. CYP2R1 is a major, but not exclusive, contributor to 25-hydroxyvitamin D production in vivo. *Proc. Natl. Acad. Sci. USA* **110**, 15650–15655 (2013).
19. Mizzi, C. *et al*. Personalized pharmacogenomics profiling using whole-genome sequencing. *Pharmacogenomics* **15**, 1223–1234 (2014).

20. Gordon, A.S. *et al*. Quantifying rare, deleterious variation in 12 human cytochrome P450 drug-metabolism genes in a large-scale exome dataset. *Hum. Mol. Genet.* **23**, 1957–1963 (2014).
21. Mandelker, D. *et al*. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet. Med.* (2016); e-pub ahead of print.
22. Twist, G.P. *et al*. Constellation: a tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene, CYP2D6, from whole-genome sequences. *NPJ Genom. Med.* **1**, article number 15007 (2016).
23. Vivot, A., Boutron, I., Ravaud, P. & Porcher, R. Guidance for pharmacogenomic biomarker testing in labels of FDA-approved drugs. *Genet. Med.* **17**, 733–738 (2015).
24. Buchard, A., Sanchez, J.J., Dalhoff, K. & Morling, N. Multiplex PCR detection of GSTM1, GSTT1, and GSTP1 gene variants: simultaneously detecting GSTM1 and GSTT1 gene copy number and the allelic status of the GSTP1 Ile105Val genetic variant. *J. Mol. Diagn.* **9**, 612–617 (2007).
25. Bush, W.S. *et al*. Genetic variation among 82 pharmacogenes: the PGRNseq data from the eMERGE network. *Clin. Pharmacol. Ther.* **100**, 160–169 (2016).
26. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
27. Adzhubei, I.A. *et al*. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
28. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. & Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
29. Wei, P., Liu, X. & Fu, Y.X. Incorporating predicted functions of nonsynonymous variants into gene-based analysis of exome sequencing data: a comparative study. *BMC Proc.* **5** (Suppl. 9), S20 (2011).
30. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
31. Capriotti, E., Fariselli, P. & Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33**, W306–W310 (2005).
32. Oda, A., Yamaotsu, N. & Hirono, S. Studies of binding modes of (S)-mephenytoin to wild types and mutants of cytochrome P450 2C19 and 2C9 using homology modeling and computational docking. *Pharm. Res.* **21**, 2270–2278 (2004).
33. Shu, Y. *et al*. Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J. Clin. Invest.* **117**, 1422–1431 (2007).
34. SEARCH Collaborative Group et al. SLCO1B1 variants and statin-induced myopathy–a genomewide study. *N. Engl. J. Med.* **359**, 789–799 (2008).
35. Sakaeda, T., Nakamura, T. & Okumura, K. Pharmacogenetics of MDR1 and its impact on the pharmacokinetics and pharmacodynamics of drugs. *Pharmacogenomics.* **4**, 397–410 (2003).
36. Loo, T.W., Bartlett, M.C. & Clarke, D.M. Arginines in the first transmembrane segment promote maturation of a P-glycoprotein processing mutant by hydrogen bond interactions with tyrosines in transmembrane segment 11. *J. Biol. Chem.* **283**, 24860–24870 (2008).
37. Rehm, H.L. *et al*. ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med.* **15**, 733–747 (2013).
38. Dunnenberger, H.M. *et al*. Preemptive clinical pharmacogenetics implementation: current programs in five US medical centers. *Annu. Rev. Pharmacol. Toxicol.* **55**, 89–106 (2015).
39. Hoffman, J.M. *et al*. PG4KDS: a model for the clinical implementation of pre-emptive pharmacogenetics. *Am. J. Med. Genet. C Semin. Med. Genet.* **166C**, 45–55 (2014).
40. Ji, Y. *et al*. Preemptive pharmacogenomic testing for precision medicine: a comprehensive analysis of five actionable pharmacogenomic genes using next-generation DNA sequencing and a customized CYP2D6 genotyping cascade. *J. Mol. Diagn.* **18**, 438–445 (2016).
41. Van Driest, S.L. *et al*. Clinically actionable genotypes among 10,000 patients with preemptive pharmacogenomic testing. *Clin. Pharmacol. Ther.* **95**, 423–431 (2014).
42. Rasmussen-Torvik, L.J. *et al*. Design and anticipated outcomes of the eMERGE-PGx project: a multicenter pilot for preemptive pharmacogenomics in electronic health record systems. *Clin. Pharmacol. Ther.* **96**, 482–489 (2014).
43. Williams, R., Peisajovich, S.G., Miller, O.J., Magdassi, S., Tawfik, D.S. & Griffiths, A.D. Amplification of complex gene libraries by emulsion PCR. *Nat. Methods* **3**, 545–550 (2006).
44. Lan, T. *et al*. Effects of CYP2C19 variants on methadone metabolism in vitro. *Drug Test. Anal.* (2016); e-pub ahead of print.
45. Wang, H. *et al*. Evaluation of the effects of 20 nonsynonymous single nucleotide polymorphisms of CYP2C19 on S-mephenytoin 4′-hydroxylation and omeprazole 5′-hydroxylation. *Drug Metab. Dispos.* **39**, 830–837 (2011).
46. Niwa, T., Kageyama, A., Kishimoto, K., Yabusaki, Y., Ishibashi, F. & Katagiri, M. Amino acid residues affecting the activities of human cytochrome P450 2C9 and 2C19. *Drug Metab. Dispos.* **30**, 931–936 (2002).