Molecular inversion probe (MIP)-based multiplexed target capture for pharmacogenomic analysis

Soo Min Han

Department of Medical Science

The Graduate School, Yonsei University

Molecular inversion probe (MIP)-based multiplexed target capture for pharmacogenomic analysis

Directed by Professor Min Goo Lee

The Doctoral Dissertation submitted to the
Department of Medical Science,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Soo Min Han

December 2015

This certifies that the Doctoral Dissertation of Soo Min Han is approved.

Thesis Supervisor: Min Goo Lee

Thesis Committee Member #1: Min Soo Park

Thesis Committee Member #2: Ji Hyun Lee

Thesis Committee Member #3: Hyun Seok Kim

Thesis Committee Member #4: Sangwoo Kim

The Graduate School

Yonsei University

December 2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Molecular inversion probe (MIP)-based multiplexed target capture for pharmacogenomics analysis

Soo Min Han

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Min Goo Lee)

As the target capture platform for the pharmacogenomics analysis, we designed a panel of molecular inversion probes to cover 80 genes which can affect exposure and response to drugs. By stepwise rebalancing, probes were optimized to capture targets with high coverage (96%) and accuracy (99.8%). The panel features the efficient probe preparation as initial microarray-based synthesis of duplex MIPs (microDuMIPs) is applied. Target capture and next-generation sequencing (NGS) of 191 individuals with this panel identified a number of unreported low-frequency genetic variants in targeted genes. *In silico* prediction showed genetic variants at low frequency enriched for ones with deleterious functional impacts. By further *in vitro* investigation, selected rare or low-frequency variants were validated to alter functions in influx/efflux transporters (SLC22A1/OCT1, SLCO1B1/OATP1B1, ABCB1/MDR1 and a metabolizing enzyme (CYP2C19). This implicates genetic variants at low frequency, which might have been overlooked in past pharmacogenetic studies, may bear relevance to the inter-individual difference in response to drugs. As a follow-up, we applied the platform to capture DNA sequences of 90 participants in phase 1 clinical trials for Tacrolimus, an

immunosuppressive drug known for large inter-individual variance in pharmacokinetics (PK). Besides CYP3A5*3, an established index variant for the PK variability, POR*28 and rare variants in UGT1A4 were revealed to be the account for the part of the PK variability. We verified that the POR*28 allele decreased exposure to drug, distinctively in CYP3A5 full expressers. Rare variants in UGT1A4 may cause variable extent of glucuronidation as phase II metabolism of secondary importance, given that this genetic effect is conspicuous in phenotypic extremes with impaired CYP3A5 function (*1/*3 and *3/*3). To the best of our knowledge, this is the first time to implicate the association of genetic variations in UGT1A4 with variable responses to Tacrolimus. Collectively, POR*28 and rare variants in UGT1A4 can explain the PK variance approximately 6% more than solely by CYP3A5*3. These results suggest that not only common but also rare variants are worth being considered for pharmacogenomic analysis and testing. Our platform also presents a proof of concept that NGS-based pharmacogenomic testing at earlier phase of clinical trials can provide the foundation for understanding metabolic pathways of drugs by extensive genetic screening including rare variants scan. Moreover, this shows that MIP-based multiplexed target capture can be one of effective tools to investigate individualized metabolic pathways of drugs.

Key words: pharmacogenomic target sequencing, ADME gene, molecular inversion probe, rare variants, tacrolimus

Molecular inversion probe (MIP)-based multiplexed target capture for pharmacogenomics analysis

Soo Min Han

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Min Goo Lee)

I. INTRODUCTION

Pharmacogenetics is to study genetic variations predisposed to phenotypic variations to drugs with the purposes of reducing unexpected adverse drug reactions (ADR) and maximizing clinical benefits of pharmacotherapies. Variants in genes affecting pharmacokinetic (PK) and pharmacodynamic (PD) factors are of great importance in pharmacogenetics. Whereas PK genes, also referred as ADME genens, are linked with drug exposure, influencing absorption, distribution, metabolism and excretion (ADME) process, PD genes encode specific drug targets and therefore, influence response to drugs directly. While previously developed genotyping platforms such as Affymetrix DMET^{TM} Plus panel cover the limited number of markers mostly single nucleotide polymorphisms (SNPs) of PK/PD genes, recent advances in sequencing technologies enable genotyping in genome-wide manner. NGS-based platforms have the virtue of the unbiased investigation, taking rare and low-frequency genetic variants into account for pharmacogenomic analysis.

1. Molecular inversion probe (MIP)

As innovative strategies of NGS, target-enrichment methods have provided the basis for various platforms and sequencing projects. Compared to multiplexing PCR and hybridization-based capture, MIP-based capture features high specificity with low amount of DNA required.[1] Besides, MIP allows various modifications on its backbone sequence for more efficient protocols as exemplified by molecular tagging.[2,3] Due to this feature, MIP used to be applied with microarrays and multi-color detection for genotyping.[4-6] On the other hand, its major drawbacks are that synthesizing MIPs at column-based is costly and often the allelic bias disrupts genotyping accuracy. Nevertheless, MIP-based capture has been evolved into various versions with many improvements and proved to fit for the large-scale sequencing of several disease conditions.[2,7-10] As one of the developed versions of MIP-based capture, we previously reported the microarray-based synthesis of duplex MIPs (microDuMIPs). It is a less costly and laborious method while keeping comparable performances by mitigating the allelic bias with unique barcoding.[3]

2. Association studies in Pharmacogenomics

Variants of genes affecting pharmacokinetic and pharmacodynamic factors contribute inter-individual variability in the phenotypes of drug response to varying degrees. Many of these variants and their functional roles have been widely reported as the fruition of genome-wide studies of drug responses or ADRs.[11,12] However, the fruitions are largely limited to common variants as yet.[13] Defining rare variations intrinsically requires the large-scale investigation, which makes a rare variant-association study harder and more costly. As implicated in sequencing data of over 10,000 samples for hundreds of drug target genes, rarer variants were predicted to be damaging and maintained at lower frequency, possibly due to the purifying selection.[14] Nevertheless, only a few studies of rare variant-association were performed for phenotypes of toxicity or drug response.[15] Therefore, contributions of rare variants to phenotypes of drug exposure and response have not yet been fully

investigated and the whole picture of roles of common and rare variants in pharmacogenomics is still veiled.

3. Pharmacogenomics in clinical practice

Pharmacogenetic studies finally aim to suggest genetic factors to be reflected in decision making at pharmacotherapies. However, clinical implementation of pharmacogenetic researches has proceeded slowly. In an effort to adopt pharmacogenetics in clinical practice, clinically important genes being paired with drugs have been published and updated through the Pharmacogenetics Knowledge Base (PharmGKB)[16] and Clinical Pharmacogenetics Implementation Consortium (CPIC).[17] Over 140 drugs have been listed with genetic biomarkers at the FDA Table of Pharmacogenomic Biomarkers in Drug Labels.[18,19] Meanwhile, a number of genetic variants with the clinical actionability are implemented in clinics,[20] but these are confined to the part of common markers. Recently, preemptive pharmacogenomic testing was suggested with proved validity,[21] but it is still controversial for its utility.[22,23]

4. Tacrolimus, an immunosuppressant of large PK variability

Tacroilmus is an immunosuppressive drug with large inter-individual variability but narrow therapeutic window. A higher level of Tacrolimus can trigger nephrotoxicity[24] and a lower level than the required concentration can cause acute rejections after organ transplantations.[25] Therefore, maintaining the optimal concentration of Tacrolimus is crucial and also related to prognosis. A canonical pharmacokinetic pathway includes two cytochrome P450 3A enzymes (CYP3A4 and CYP3A5) for metabolizing and ABCB1/MDR1 for an efflux transporter. Importantly, CYP3A5 plays predominant roles in oxidative metabolism such as O-demethylation and hydroxylation.[26] After the existence of conjugated Tacrolimus to glucuronide was reported in human bile,[27] uridine 5'-diphosphate glucuronosyltransferase (UGT) 2B7 and 1A4 had been suggested for major

enzymes for glucuronidation.[28,29] However, no significant mechanistic researches on phase II metabolism of Tacrolimus has been performed ever since initial discoveries. Tacrolimus is classified in calcineurin inhibitors, of which mechanism of action includes inhibiting the phosphatase activity of calcineurin. It is known that forming a complex with FKBP12 renders the increase of the binding affinity of Tacrolimus to calcineurin.[30] Inhibition of calcinuerin activity, in turn, deactivates the immune response in T lymphocytes.[31,32] According to PharmGKB, previous studies have suggested several genetic markers in three canonical genes and other genes such as POR, NR1I2, IL10, IL18, PPARA and TLR4 for the PK variance. However, only CYP3A5*3 could be introduced as the established genetic marker for a dosing guideline of Tacrolimus.[33] Other markers have been only suggested with weak evidences due to inconsistent results on different cohorts.

5. Dissertation aims

Here, by applying the microDuMIP technique, we primarily aim to generate the optimized MIP set with high coverage and accuracy to capture targeted PK/PD genes. The optimized MIP set is to be applied for 191 healthy Koreans. Through high-throughput sequencing data, we attempt to investigate the distribution of unreported rare or low-frequency variants with reference to databases of population genetics. With selected variants being appeared in low frequency, we aim to suggest potential roles of those variants in influx/efflux transporters and metabolizing enzymes by evaluating functional effects *in silico* and *in vitro*. Aside from the functional validation, we utilize the MIP set to healthy participants for two phase 1 clinical trials of Tacrolimus. We not only attempt to investigate genetic determinants of variability in representative PK parameters but also aim to add the explanatory information of genetics regarding the PK variability in Tacrolimus-treated individuals by exploring unknown genetic associations of novel or rare variants. The aim and progress of the study is depicted (Figure 1).

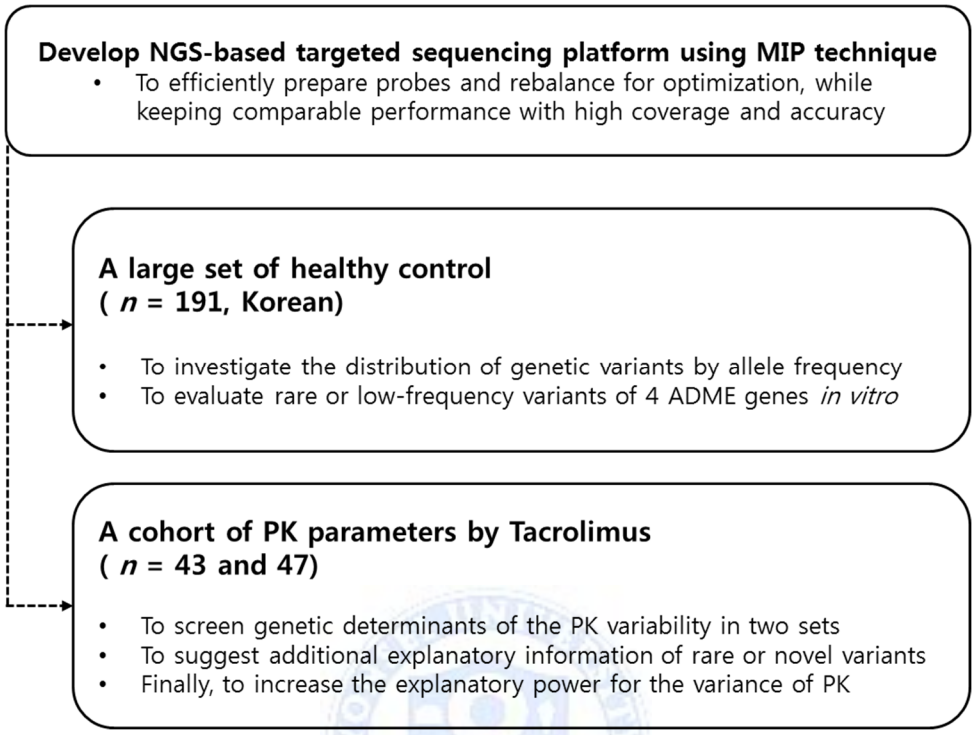

Figure 1. The study flow diagram and objectives at each phase.

II. MATERIALS AND METHODS

1. Design, generation and optimization of MIPs

To synthesize probes on a microarray, probe sequences were designed by fetching desirable lengths of annealing sequences from the reference genome sequence (hg 19/Build 37). It is known that most efficient capture is performed by optimal melting temperatures(Tm) of two annealing parts of probes. Based on the preliminary data that suggested an optimal range of Tm, annealing sequences were determined while keeping each side of adjacent probes to be overlapped. Each probe consisted of 166 bases, of which sequence was constructed with annealing and flanking sequences, common linker, and randomly distributed 15 bases as molecular tagging (Figure 2). Oligonucleotides were synthesized at a microchip-base and amplified to get a sufficient amount of probes for capturing target genes from hundreds of genomes. For an even and efficient MIP amplification, emulsion PCR was adopted with by slight modifications of the previous protocol.[34] Mixture of water-in-oil (w/o) emulsion was uniformly circulated by using ULTRA-TURRAX® Tube Drive Control (IKA® Works, Inc., Wilmington, NC, USA). The 100 μl aliquots of emulsion from the PCR mixture was amplified using the following PCR condition: 95°C for 3 min, 35 cycles of 20 sec at 95°C, 30 sec at 58°C and 30 sec at 72°C, and 10 min at 72°C. PCR-amplified samples were gel-loaded and purified at the correctly sized bands of 166 bp with the gel extraction kit (QIAGEN, Valencia, CA, USA). Subsequent PCR was performed with 1 μl of previously gel-purified product and its PCR condition was as same as above but for 25 cycles. 17 μl aliquots of gel-extracted product was digested for 16 hr at 37℃ by 1 μl of *Ear*I (New England Biolabs, Ipswich, MA, USA) with 2 μl of CutSmartTM buffer (New England Biolabs, Ipswich, MA, USA) to cleave flanking sequences used for PCR amplification. By the gel electrophoresis, enzyme-digested products were excised at the correct size and those from 4 lanes of a gel were purified by one column provided by gel extraction kit (QIAGEN, Valencia, CA, USA).

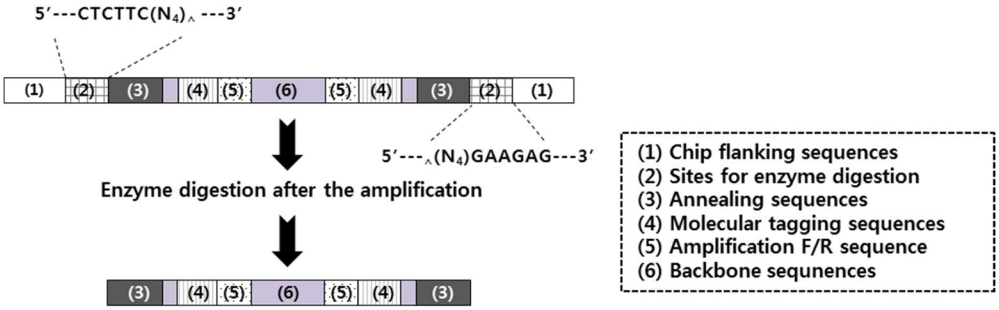

Figure 2. The design of a MIP sequence. (1) Common chip flanking sequences to amplify probes after the cleavage from a microarray. (2) Enzyme sites for *Ear*I to cleave the flanking sequences after amplification. (3) Sequences to anneal to both sides of target regions. (4) Molecular tagging for unique barcodes ranged at 15 bases, composed of randomly designated 4 bases. (5) Forward and reverse sequences to amplify from hybridized probes for libraries. (6) Backbone sequences to support the length of probes.

2. Sample collection

Genomic DNAs of total 281 healthy individuals were collected from clinical pharmacology trials performed at Seoul National University Hospital. Among those, 191 individuals were accompanied with genotypes by the DMET plus panel. We obtained two sets of 43 and 47 individuals, who were participated in two independent phase 1 clinical trials of oral taking of one or two capsules of 1mg Tacrolimus as a single dose. All participants were provided with written informed consents before their samples were collected.

3. Target capture and high throughput sequencing

Total 1 μg of genomic DNA for each individual and amplified MIPs at 1:100 ratio were mixed with Ampligase buffer (Epicentre, Madison, WI, USA) and dH2O to make a total volume of 15 μl. Hybridization started with denaturing for 2 min at 94°C, being ramped at 0.1°C per second to 60°C and incubated for 48 hr at 60°C. Then, 2 U of AmpliTaq® DNA polymerase (LifeTechnologies, Carlsbad, CA, USA),

4U of Ampligase DNA ligase (Epicentre, Madison, WI, USA), 10X dNTPs (New England Biolabs, Ipswich, MA, USA), 0.2 μL of Ampligase buffer (Epicentre, Madison, WI, USA) were added and the mixtures were incubated for 24 hr at 60°C. Next, 0.5 μL of Exonuclease I (New England Biolabs, Ipswich, MA, USA) and 0.5 μL of Exonuclease III (New England Biolabs, Ipswich, MA, USA) were used to remove linear DNA fragments at 37°C for 2 hr, and the incubation at 94°C for 4 min was followed for the deactivation. To amplify and attach barcodes and indexes for NGS, 1 μL of a hybridized template, 10 μL of KAPA HiFi polymerase (KAPA biosystems, Wilmington, MA, USA), 7 μL of dH2O with 1 μL each of AmpF (5'-CAG ATG TTA TCG AGG TCC GAC-3') and AmpR (5'- GGA ACG ATG AGC CTC CAA C-3') primers were mixed and amplified by PCR. Prepared libraries were validated for quality control with Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and sequenced further by the Illumina HiSeq2500 platform according to manufacturer's instructions.

4. Bioinformatics and in-house programs for genotype calling

Raw sequencing data were separated by sample-specific barcodes. The annealing portions (up to 30bp each) in pair-end reads were removed and 15 bases of molecular tags were preserved in headers of fastq formats (Figure 3). Only gap-filled sequences were mapped to the reference genome sequence (hg 19/Build 37) using Novoalign (V2.07.18). Local realignments near in-dels were performed and quality scores were recalibrated by using the Genome Analysis Toolkit (v2.3.6). PCR duplicates were discarded by unique molecular tag information through the in-house program (Figure 3), which resembles Picard (v1.6.7), a commonly used program for NGS data analysis. GATK UnifiedGenotyper (v2.3.6) was used for genotype calling for each sample separately. Format converting, sorting, and indexing were performed by Samtools(v.0.1.19). Annotation and functional effect prediction were performed by PolyPhen-2 (v2.2.2), SIFT (v.1.03) and ANNOVAR.

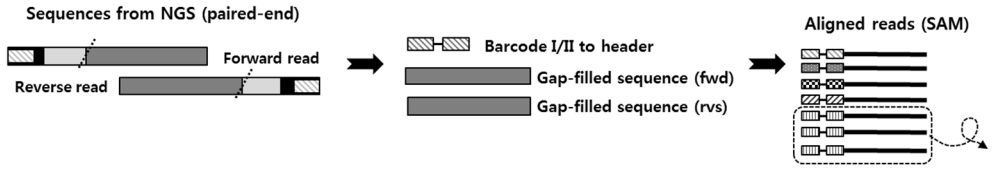

Figure 3. In-house program for trimming barcodes and removing duplicates. The barcode information from forward and reverse reads were together preserved in headers of fastq formats to remove duplicates from aligned sequencing reads. This helps adjusting the allelic bias due to PCR duplicates.

5. Performance test and variant selection for *in vitro* assays

The F1 score, defined as 2·[(precision·recall)/(precision + recall)], was used for setting empirical filters for heterozygous genotype callings. After filtering genotype calls with low confidence (sequencing depth ≤ 15 and variant allele frequency ≤ 0.3), genotyping by the optimized MIP set was compared with benchmark genotype calls for NA12878.[35] Total 485 of SNP and in-del genotypes by our platform were compared with those by the commercial genotype platform, DMET plus panel (*n*=1927).

Genetic variants that met the following criteria were collected for *in vitro* assay. 1) Nonsynonymous variants of *SLC22A1, ABCB1, SLCO1B1* and *CYP2C19*. 2) Rare variants in the general population, which were discovered only in one or two individuals in 191 healthy people and never appeared in databases of dbSNP and 1000genome project (1000GP).[36] 3) Deleterious consequences of variants expected by at least one functional prediction programs (SIFT[37] or polyphen-2[38]).

6. Site-directed mutagenesis, cell culture and transfection

The expression clones of OATP1B1, OCT1 and CYP2C19 were given by Dr Sang Seop Lee (Inje University). For mutagenesis, PCR was performed on a template vector by using two primers that have a mismatched base at the site where to be mutated. Pfu-X DNA polymerase (Solgent, South Korea) with the supplied

10X reaction buffer and dNTP mix (10 mM) were used. PCR amplification consisted of initial denaturation at 95°C, followed by 30 cycles of denaturation at 94°C, annealing at 68°C, and elongation at 68°C for 5 min. The PCR product was digested by *Dpn*I for 2 hr at 37°C and finally transformed into DH5α *E. coli* cells (RBC Bioscience, New Taipei City, Taiwan). Mutant plasmids were isolated using the Plasmid DNA Mini-Prep kit (Intron, South Korea) from ampicillin-resistant colonies. All mutations were confirmed by sequencing the full-length of all mutant clones generated. HEK 293T cells or HEK 293 cells were maintained in Dulbecco's Modified Eagle Medium (Invitrogen, Carlsbad, CA, USA) containing 10% fetal bovine serum at 37℃. Plasmids were transiently transfected by using Lipofectamine Plus Reagent or Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA).

7. Calcein AM efflux assay for ABCB1/MDR1

Two days after transfection with pcDNA3.1(+)-hMDR1 expression vectors, HEK 293T cells were dispersed by trypsinization and washed twice with ice cold DMEM supplemented with 2 mM L-glutamine and 1 mM sodium pyruvate (Invitrogen, Carlsbad, CA, USA). To allow Calcein AM accumulation, the total volume of 500 µl of 1 x 10^6 cells was incubated for 60 min at 37℃ with the medium containing 1 µM Calcein AM (Abcam, San Francisco, CA, USA) with or without 1 µM Elacridar (Santa Cruz Biotechnology, Dallas, TX, USA), a MDR1 specific inhibitor. Next, cells were analyzed immediately by a FACSVerse flow cytometer (BD Bioscience, San Jose, CA, USA). The median fluorescence of Calcein (ex/em 588/527nm) was determined and compared.

8. [3H]-Esteron-3-Sulfate (ES) uptake assay for SLCO1B1/ OATP1B1

Cells transfected in 24-well plates were washed twice with 500 µl of pre-warmed 37℃ DPBS and incubated in 37℃ water bath for 10 min. After removing DPBS in each well, 300 µl of pre-warmed DPBS with [3H]-ES was added and

incubated for 5 min for the uptake. Each well was washed with ice-cold DPBS for 3 times and cells were solubilized with 200 µl of 0.1N NaOH by shaking for an hour. Once solubilized, cells were transferred to a new 23-well plate with 500 µl of OptiPhase SuperMix (PerkinElmer, Waltham, MA, USA) and shaken for 4 hr. The radioactivity of [3H]-ES in cell lysates was measured with Beta counter.

9. [14C]-tetraethylammonium(TEA) uptake assay for SLC22A1/OCT1

pcDNA3.1-hOCT2 wild-type and mutants were transiently transfected to HEK 293 cells. Two days after transfection, cells were washed with DMEM and maintained for an hour with serum-free DMEM at 37°C. The washing medium was replaced with 1 ml of serum-free DMEM containing 100 µM of [14C]-TEA to allow the uptake. The uptake was quenched by washing with 2 ml of ice-cold phosphate-buffered saline for 3 times. The cells were lysed with 100 µl of cell lysis reagent (Promega, Madison, WI, USA). Then, the radioactivity was measured.

10. Purification and expression of CYP2C19

DH5α *E. coli* competent cells were transformed with pCW vectors of *CYP2C19* wild-type and mutated type. A single colony was inoculated to 100 ml LB containing 100 µg/ml ampicillin at 37℃. Each of 50 ml overnight culture was transferred to 450 ml TB media with 100µg/ml ampicillin and kept shaken for 3 hr until the optical density reached the range within 0.4-0.6. After cooling the media for an hour, isopropyl β-D-thiogalactopyranoside (IPTG, 0.5mM at final volume) and δ-aminolevulinic acid (δ-ALA, 0.5 mM at final volume) were added and incubated at 22℃ while shaken at 150 rpm for 3 days. After we confirmed CYP2C19 expression by measuring Co difference spectrometry, the cultures were centrifuged at 5000*g* for 10 min at 4℃. The pellets were resuspended with 15 ml of 2X TES/g cells until no clumps were detected at 4℃. After adding 0.6 ml lysozyme, 15 ml of cold DW was added immediately and incubated for 30 min on ice while

stirrred. The suspension was centrifuged at 5000-10000*g* for 10 min at 4℃ and supernatants were removed. The sonication buffer of 2 ml/g (wet cells) with DTT was added and resuspended with gentle mixing. Then, the mixture was further sonicated and centrifuged at 10000*g* for 20 min at 4℃. The supernatants were centrifuged at 150000*g* for 2 hr and the isolated membranes were resuspended with 10 ml of 1X TES buffer assuming 1 ml/g (wet cells).

11. 4'-hydroxylation assay of S-mephenytoin by CYP2C19

We reconstituted CYP2C19 protein (5 pmol) with cytochrome b5 (10 pmol), 0.05 μmol sodium cholate and human NADPH-P450 oxidoreductase (20 pmol) in TES buffer (pH 7.4). The hydroxylation reaction was initiated by incubating with 10 mM NADPH at 37℃. For the reaction, 200 μM of S-mephenytoin was used. The enzyme activity for 4'-hydroxy S-mephenytoin was quantified by a Qtrap 4000 liquid chromatography/tandem mass spectrometry system (Applied Biosystems, Foster City, CA, USA) equipped with electrospray ionization.

12. Statistical analysis

For results of *in vitro* investigation, data was presented as mean ± SEM for at least three experiments. Statistical significance was marked from Student's *t*-tests, at which $P < 0.05$ was regarded as statistically significant. For comparison, unpaired *T*-test was performed for two groups with equal variances from *F*-test. Welch's corrected *T*-test was performed for two groups with unequal variances from *F*-test. Linkage disequilibrium(LD) was analyzed using Haploview software[39] within 500kb from an index locus. For rare variants of allele frequency at 0.01, simple gene-based burden tests were performed for all samples and 3 metabolizers groups categorized by CYP3A5*3 genotypes. Initially, we compared means of carrier and non-carrier of rare variants in each gene with unpaired T-test within pooled samples (n=90). Among PK parameters, we used the area under the blood concentration time

curve at last quantifiable concentration, AUC_{last}, for an index measure of drug exposure. Dose-adjusted AUC_{last} was used for the analysis of pooled samples simply by dividing AUC_{last} values by the dose. For genes with significant *P*-values, we further investigated effects of rare variants in each set (n=43 and 47). Multiple linear regression was used to calculate R^2 by adding genetic determinants including CYP3A5*3 genotypes. The significance of adding new genetic factors was checked using ANOVA, comparing the model before and after adding new genotypes. We defined the difference of R^2 as the explanatory information supplemented by each genetic factor.

III. RESULTS

1. Targeted capture platform was generated based on MIP technique for 80 genes on drug response and exposure

We initially designed 10908 probes to capture exons, promoter regions, and splicing junctions of 80 genes, covering all transcript isoforms reported (Table 1A, Figure 4A). Specifically, 600kb of upstream, +50bp/-10bp of each splicing junctions and 100kb of downstream for each gene were included as target regions. We also designed probes to target markers of the DMET Plus array including single and multi-nucleotide polymorphisms, insertions and deletions (Table 1B, Figure 4B). Two probes per each target loci were designed to prevent missing targets. We adopted the microDuMIP method with minor modifications.[40] Molecular tags of 15 bases were designated at two each sides of the probe sequence with 9 bases and 6 bases and preserved to NGS data (Figure 4C). These barcodes were introduced to allow the number of distinct molecular tags to be greater than total number of probes at the time of the synthesis. Optimization of MIPs was repeated for improving overall capture performances (Figure 5A). It is known that the adequate Tm is one of valuable predictors of the capture efficiency as it is amendable when designed.[7] Annealing sequences for MIPs were fetched from the reference genome sequence while lengths and locations were adjusted (Figure 5B). The length of arm sequences ranged from 15 bp to 27 bp. Next, the ideal range of Tm for two annealing sequences was empirically defined at the pilot stage with 16 samples (Figure 5C). Most efficient Tm ranged from 55 to 62.5℃. Interestingly, the capture efficiency was relatively well preserved even for probes with only one annealing sequence, of which Tm fit within the best range. This suggested that only one annealing portion with the sufficient capture performance was enough to support another annealing part to better hybridize to target sequences. After we set less efficient probes by the mean sequencing depth over pilot samples, we rebalanced those to fit to the ideal range of Tm by adjusting the length of annealing sequences

flexibly or shifting the location from GC- or AT-rich regions. We increased the amount of probes with low capture performance despite the ideal range of Tm. Two independently generated probe sets of the original and rebalanced were mixed at ratio 1:1, which showed better capture performances in a breadth of coverage (%) over target regions (Figure 6A). We confirmed that there was no remarkable competition between the original and rebalanced probes that blocked each other from capturing target sequences. Overall capture performance of MIPs was improved and the proportion of poorly performing MIPs decreased after rebalancing (Figure 6B).

Table 1. 80 genes and markers targeted by the MIP panel

A. 80 Target genes		
Class	Gene	Full name
Phase I[1]	*ADH1A*	Alcohol dehydrogenase 1A (class I), alpha polypeptide
Phase I	*ADH1B*	Alcohol dehydrogenase 1B (class I), beta polypeptide
Phase I	*ADH1C*	Alcohol dehydrogenase 1C (class I), gamma polypeptide
Phase I	*ALDH1A1*	Aldehyde dehydrogenase 1 family, member A1
Phase I	*CYP1A1*	Cytochrome P450, family 1, subfamily A, polypeptide 1
Phase I	*CYP1A2*	Cytochrome P450, family 1, subfamily A, polypeptide 2
Phase I	*CYP2A6*	Cytochrome P450, family 2, subfamily A, polypeptide 6
Phase I	*CYP2B6*	Cytochrome P450, family 2, subfamily B, polypeptide 6
Phase I	*CYP2C19*	Cytochrome P450, family 2, subfamily C, polypeptide 19
Phase I	*CYP2C8*	Cytochrome P450, family 2, subfamily C, polypeptide 8
Phase I	*CYP2C9*	Cytochrome P450, family 2, subfamily C, polypeptide 9
Phase I	*CYP2D6*	Cytochrome P450, family 2, subfamily D, polypeptide 6
Phase I	*CYP2E1*	Cytochrome P450, family 2, subfamily E, polypeptide 1
Phase I	*CYP2J2*	Cytochrome P450, family 2, subfamily J, polypeptide 2
Phase I	*CYP3A4*	Cytochrome P450, family 3, subfamily A, polypeptide 4
Phase I	*CYP3A5*	Cytochrome P450, family 3, subfamily A, polypeptide 5
Phase I	*CYP4F2*	Cytochrome P450, family 4, subfamily F, polypeptide 2
Phase I	*DPYD*	Dihydropyrimidine dehydrogenase
Phase I	*EPHX1*	Epoxide hydrolase 1, microsomal (xenobiotic)
Phase II[2]	*GSTM1*	Glutathione S-transferase mu 1
Phase II	*GSTP1*	Glutathione S-transferase pi 1
Phase II	*NAT1*	N-acetyltransferase 1 (arylamine N-acetyltransferase)
Phase II	*NAT2*	N-acetyltransferase 2 (arylamine N-acetyltransferase)
Phase II	*SULT1A1*	Sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1
Phase II	*TPMT*	Thiopurine S-methyltransferase
Phase II	*UGT1A1*	UDP glucuronosyltransferase 1 family, polypeptide A1
Phase II	*UGT1A10*	UDP glucuronosyltransferase 1 family, polypeptide A10
Phase II	*UGT1A3*	UDP glucuronosyltransferase 1 family, polypeptide A3
Phase II	*UGT1A4*	UDP glucuronosyltransferase 1 family, polypeptide A4
Phase II	*UGT1A5*	UDP glucuronosyltransferase 1 family, polypeptide A5
Phase II	*UGT1A6*	UDP glucuronosyltransferase 1 family, polypeptide A6
Phase II	*UGT1A7*	UDP glucuronosyltransferase 1 family, polypeptide A7
Phase II	*UGT1A8*	UDP glucuronosyltransferase 1 family, polypeptide A8
Phase II	*UGT1A9*	UDP glucuronosyltransferase 1 family, polypeptide A9
Phase II	*UGT2B15*	UDP glucuronosyltransferase 2 family, polypeptide B15

Phase II	*UGT2B7*	UDP glucuronosyltransferase 2 family, polypeptide B7
Transporter	*ABCA1*	ATP-binding cassette, sub-family A (ABC1), member 1
Transporter	*ABCB1*	ATP-binding cassette, sub-family B (MDR/TAP), member 1
Transporter	*ABCB11*	ATP-binding cassette, sub-family B (MDR/TAP), member 11
Transporter	*ABCC2*	ATP-binding cassette, sub-family C (CFTR/MRP), member 2
Transporter	*ABCC3*	ATP-binding cassette, sub-family C (CFTR/MRP), member 3
Transporter	*ABCC4*	ATP-binding cassette, sub-family C (CFTR/MRP), member 4
Transporter	*ABCC7*	ATP-binding cassette, sub-family C (CFTR/MRP), member 7
Transporter	*ABCG1*	ATP-binding cassette, sub-family G (WHITE), member 1
Transporter	*ABCG2*	ATP-binding cassette, sub-family G (WHITE), member 2
Transporter	*SLC10A1*	Solute carrier family 10 (sodium/bile acid cotransporter family), member 1
Transporter	*SLC15A1*	Solute carrier family 15 (oligopeptide transporter), member 1
Transporter	*SLC15A2*	Solute carrier family 15 (H+/peptide transporter), member 2
Transporter	*SLC22A1*	Solute carrier family 22 (organic cation transporter), member 1
Transporter	*SLC22A11*	Solute carrier family 22 (organic anion/urate transporter), member 11
Transporter	*SLC22A12*	Solute carrier family 22 (organic anion/urate transporter), member 12
Transporter	*SLC22A2*	Solute carrier family 22 (organic cation transporter), member 2
Transporter	*SLC22A3*	Solute carrier family 22 (extraneuronal monoamine transporter), member 3
Transporter	*SLC22A4*	Solute carrier family 22 (organic cation/ergothioneine transporter), member 4
Transporter	*SLC22A5*	Solute carrier family 22 (organic cation/carnitine transporter), member 5
Transporter	*SLC22A6*	Solute carrier family 22 (organic anion transporter), member 6
Transporter	*SLC22A8*	Solute carrier family 22 (organic anion transporter), member 8
Transporter	*SLCO1A2*	Solute carrier organic anion transporter family, member 1A2
Transporter	*SLCO1B1*	Solute carrier organic anion transporter family, member 1B1
Transporter	*SLCO1B3*	Solute carrier organic anion transporter family, member 1B3
Transporter	*SLCO2B1*	Solute carrier organic anion transporter family, member 2B1
PD	*ACE*	Angiotensin I converting enzyme (peptidyl-dipeptidase A) 1
PD	*ADRB2*	Adrenergic, beta-2-, receptor, surface
PD	*BRCA1*	Breast cancer 1, early onset
PD	*COMT*	Catechol-O-methyltransferase
PD	*DRD2*	Dopamine receptor D2
PD	*F5*	Coagulation factor V (proaccelerin, labile factor)
PD	*G6PD*	Glucose-6-phosphate dehydrogenase
PD	*HMGCR*	3-hydroxy-3-methylglutaryl-Coenzyme A reductase
PD	*MTHFR*	5,10-methylenetetrahydrofolate reductase (NADPH)

PD	*NQO1*	NAD(P)H dehydrogenase, quinone 1
PD	*P2RY1*	Purinergic receptor P2Y, G-protein coupled, 1
PD	*P2RY12*	Purinergic receptor P2Y, G-protein coupled, 12
PD	*PTGIS*	Prostaglandin I2 (prostacyclin) synthase
PD	*SCN5A*	Sodium channel, voltage-gated, type V, alpha subunit
PD	*TYMS*	Thymidylate synthetase
PD	*VDR*	Vitamin D (1,25- dihydroxyvitamin D3) receptor
PD	*VKORC1*	Vitamin K epoxide reductase complex, subunit 1
Modifier	*AHR*	Aryl hydrocarbon receptor
Modifier	*NR1I2*	Nuclear receptor subfamily 1, group I, member 2

B. Affymetrix DMET™ Plus panel
1928 markers of 231 genes (1842 SNPs, 2 DNPs, 1 TNP, 39 insertions, 42 deletions, 1 in-del, 1 STR)

(A) 80 genes are comprised of genes encoding Phase I and II metabolizing enzymes,[1,2] drug transporters and other genes related to drug response. (B) It also target known markers by Affymetrix's DMET Plus panel. DNP, Double Nucleotides polymorphism; TNP, Triple Nucleotides polymorphism; STR, Short Tandem Repeats.

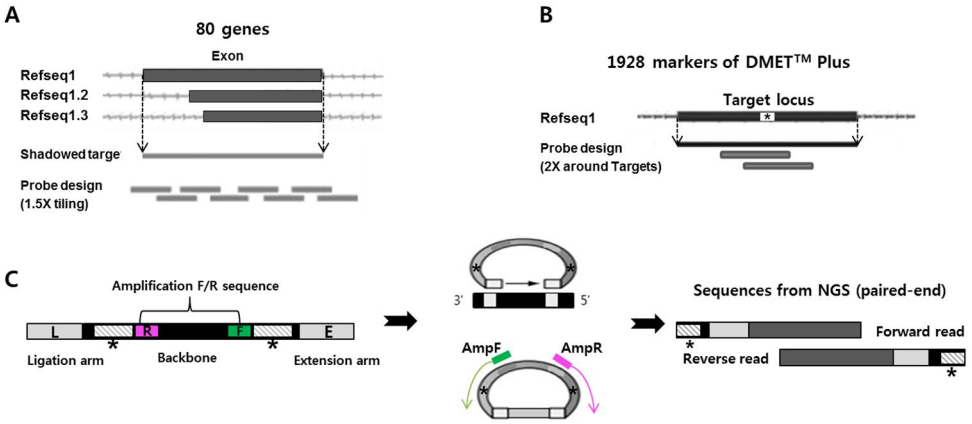

Figure 4. Molecular inversion probes (MIP) design and molecular tagging with barcodes. The design scheme of MIPs to capture exons, promoter regions, and splicing junctions of all transcript isoforms of 80 genes (A) and markers of Affymetrix's DMET plus array (B). (C) Molecular barcode tagging of randomly designated 15 bases, preserved to paired-end sequences from NGS. AmpF/R sequences were used to amplify gap-filled probes for NGS libraries.

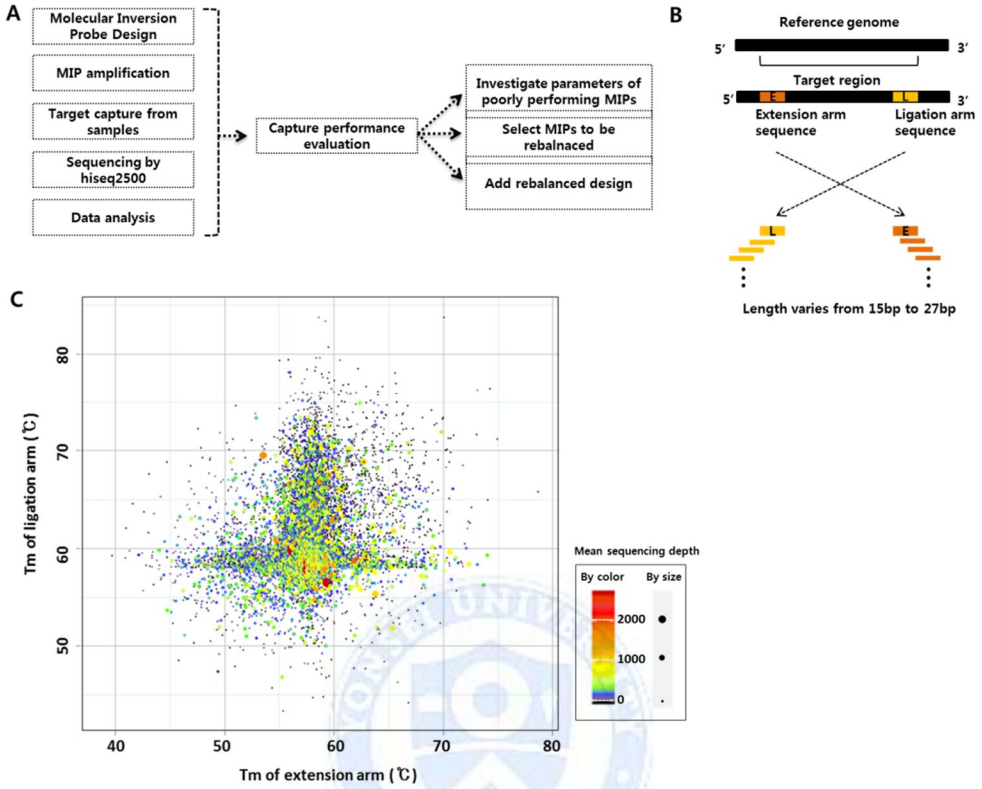

Figure 5. Optimization of MIPs and parameters for the capture efficiency. (A) The flowchart of stepwise rebalancing. The process was repeated until the improvement in the capture performance was confirmed. (B) The diagram of fetching two annealing arm sequences from the reference sequence. Variable lengths and locations were considered to fit in the best Tm range. (C) The predictor of the capture efficiency for 10908 MIPs. The mean sequencing depth over the target region by each MIP was calculated and regarded as an index of the capture efficiency. These were color-coded and scaled by dots in the plot.

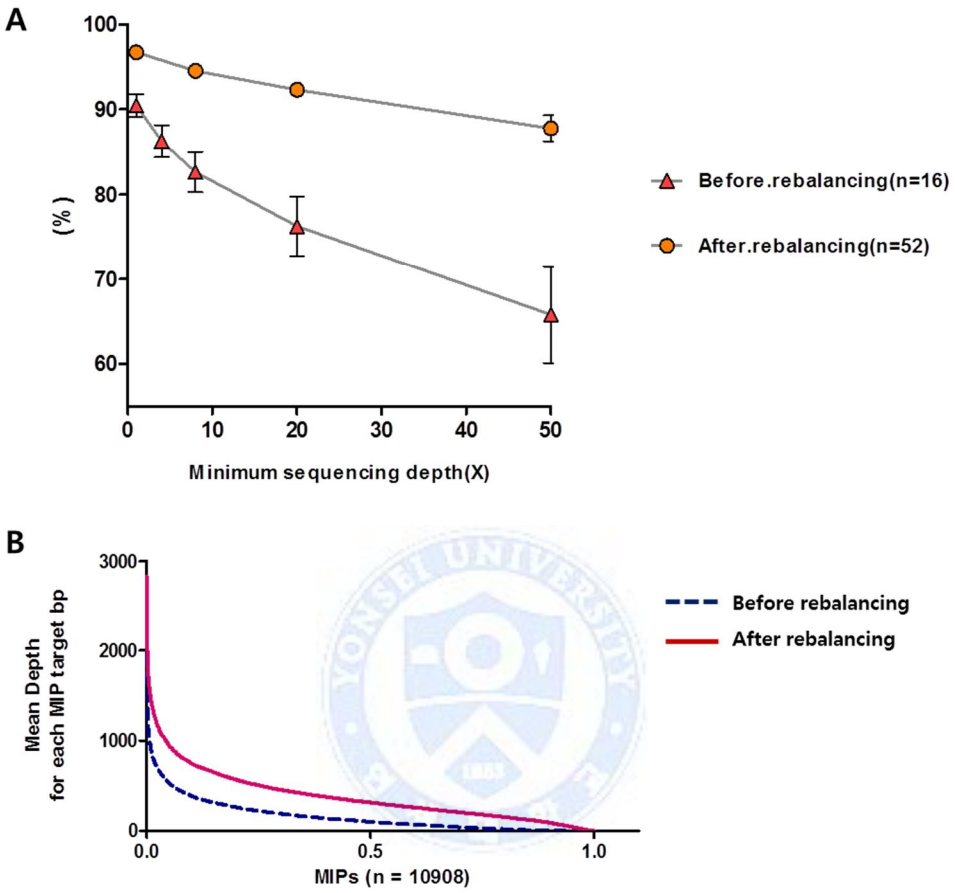

Figure 6. Performance comparison of MIPs after rebalancing. (A) Overall improvement of a breadth of coverage (%) over target regions before and after rebalancing MIPs. (B) MIPs with low performance were improved to capture targets with higher coverage. The proportion of poorly performing MIPs decreased after rebalancing.

2. Optimized MIPs enable efficient sequencing of ADME and PD genes with high coverage and accuracy

The optimized MIP set was applied to capture target sequences from 191 healthy Koreans (Figure 7). Exploring the breadth of coverage over target regions, median 96% of targets were captured and 92% at 15X sequencing depth. Covered regions were sequenced to the median depth of 400X, showing the uniform distribution of coverage. The MIP sets comprised of two kinds of probes, which were exon-targeting MIPs for 80 genes and locus-targeting MIPs for markers of DMET plus panel. For locus-targeting MIPs, median 98% of targets were captured and 96% at 15X sequencing depth.

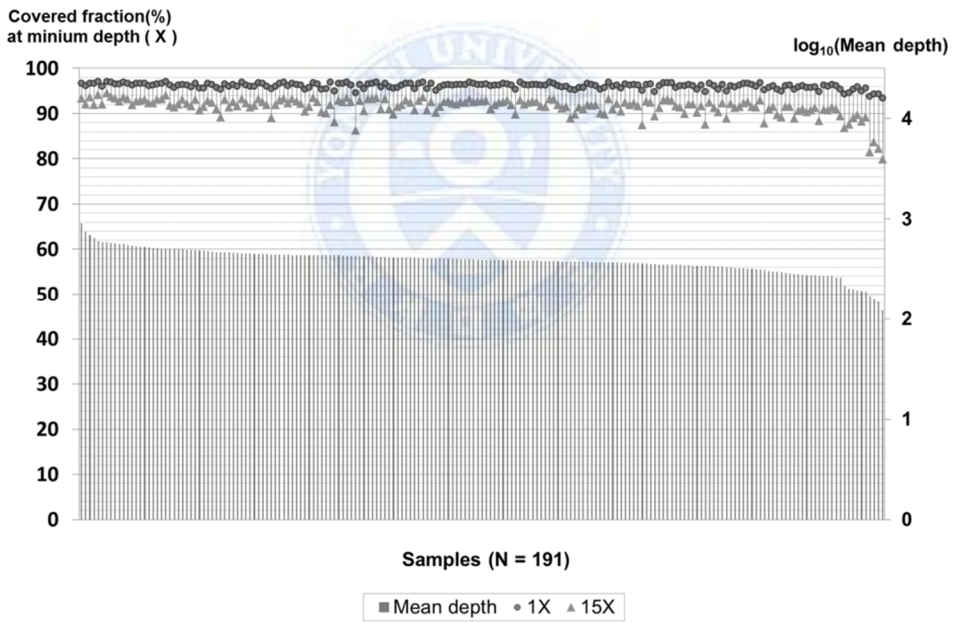

Figure 7. The summary statistics of sequencing data for 191 Koreans. The breadth of coverage was depicted at the left-Y-axis and mean depth at the right-Y-axis. Median 96% of targets were covered and 92% at 15X sequencing depth with median NGS depth of 400X.

We set the empirical filter for the genotype calling, especially for heterozygous calls. By adjusting the filter conditions for sequencing depth and allelic depth, the F1 score was calculated for genotype comparisons between the MIP set and validated genotypes of NA12878. We defined the filter as sequencing depth ≥ 15 and variant allele frequency ≥ 0.3 (Figure 8). After filtering genotype calls with low confidence, genotypes by the MIPs were compared with benchmark genotype calls, validated by integrating multiple sequencing data.[35] Total 485 of SNP and in-del genotypes were 99.8% concordant with the validated genotypes with 98.1% sensitivity (Table 2). When compared to using the commercial genotype platform, DMET plus panel, genotypes by our platform showed 99.4% concordance with 97% sensitivity (Table 2).

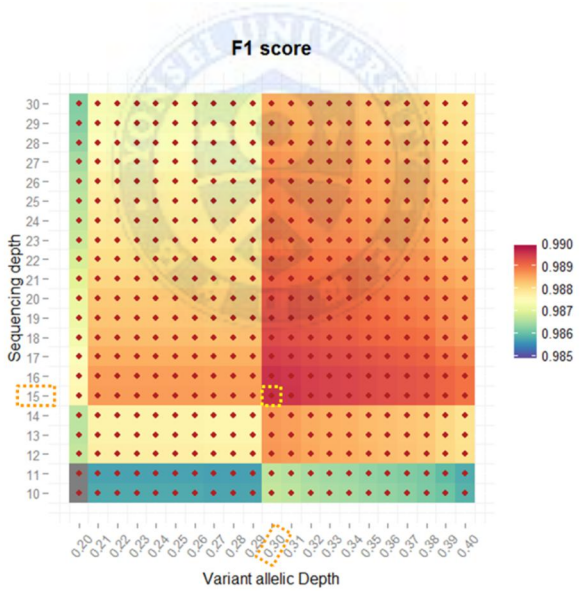

Figure 8. F1 Scores as adjusting filtering conditions. As variant allelic depth (X-axis) and sequencing depth (Y-axis) were adjusted, the empirical filtering condition was set by measuring the F1 score, which represents PPV and recall (sensitivity).

Table 2. Comparison of genotyping by the optimized MIP set and existing platforms

	Affymetrix DMET^{TM} Plus panel	Validated genotype for NA12878
Platform	Array	WES, WGS
Number of sites common	1927	485
Number of Samples compared	191	1 (NA12878)
PPV (%)	99.4 (mean)	99.8
Sensitivity (%)	97 (mean)	98.1

Positive Predictive Value (PPV, %) and Sensitivity (%) of genotypes by the optimized MIP set, when comparing to those from DMET plus panel and integrated multiple data. For comparison with DMET plus panel, the mean of 191 samples were described in the table.

3. Targeted sequencing of ADME genes for healthy Koreans reveals many unidentified functional variants at low allele frequency

With variant profiles for 191 healthy Koreans, we explored the distribution of nonsynonymous and synonymous variants by their allele frequencies. The ratio of nonsynonymous to synonymous variants (NS:S ratio) had a tendency to increase as variants were detected in lower frequency (Figure 9A). In fact, the number of samples (n=191) was not enough to detect very rare variants. Nevertheless, the tendency indicated that many of function-altering variants have been under the purifying selection, which made those variants kept in low frequency in the given population. This was consistent with the previous results of sequencing 202 drug target genes over ten thousands of people.[14] By *in silico* functional prediction programs (SIFT[37] and Polyphen-2[38]), variants at low allele frequency were more likely to have damaging impacts on functions of proteins (Figure 9B). The result suggested that not all but many of genetic variants in low frequency may have damaging effects on protein functions. In pharmacogenetics, it suggested that genetic variants in ADME or PD genes, even at low frequency, can have substantial effects on response and exposure to drugs.

In an attempt to evaluate functional consequences, unprecedented variants, which were observed only once or twice in 191 samples, were selected (Table 3). Due the fact that the number of samples was not sufficient to detect rare variants, we checked databases of population genetics such as dbSNP and 1000GP. Considering that more than thousands of samples were involved in each database, target variants for functional evaluation were selected if never reported in public databases. Reported variants with different variant alleles were regarded as novel. Variants with predicted damaging effects were selected preferably in influx/efflux transporters (SLC22A1/OCT1, SLCO1B1/OATP1B1 and ABCB1/MDR1) and metabolizing enzyme (CYP2C19).

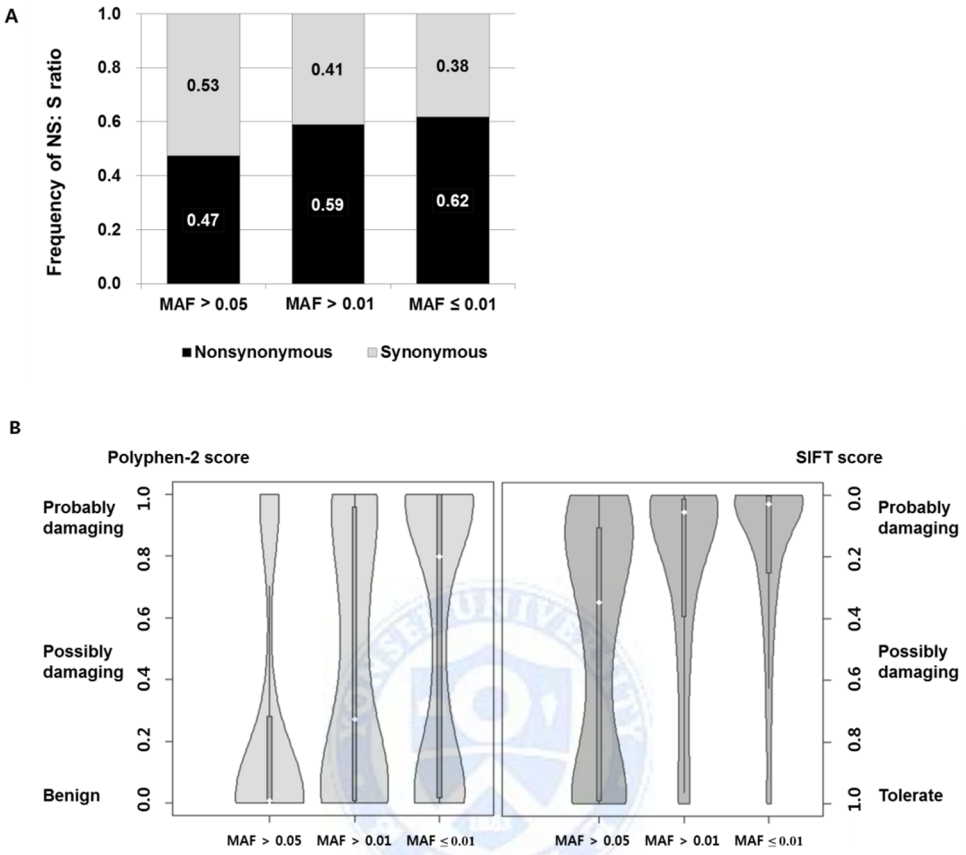

Figure 9. Properties of genetic variants by allele frequency (AF) observed in 191 Koreans. (A) The portion of nonsynonymous variants was elevated at low allele frequency. (B) Variants at lower frequency were predicted to be functionally deleterious by two *in silico* functional prediction programs (Polyphen-2 and SIFT). MAF, Minor Allele Frequency.

Table 3. Unreported variants at low allele frequency for *in vitro* functional validation

Gene	Variant	Protein [1]	Frequency	pph-2	SIFT	Novelty [2]
SLC22A1	c.C98T	p.A33V	2/191	Possibly damaging	Damaging	Novel
SLC22A1	c.A272G	p.Y91C	1/191	Probably damaging	Damaging	Novel
SLC22A1	c.G1219A	p.A407T	1/191	Probably damaging	Damaging	Novel
SLCO1B1	c.G250T	p.V84L	2/191	Probably damaging	Damaging	Novel
SLCO1B1	c.G1213T	p.V405F	1/191	Benign	Tolerated	Different allele [3]
ABCB1	c.G751A	p.G251R	1/191	Probably damaging	Damaging	Different Allele [4]
ABCB1	c.C2278G	p.L760V	1/191	Possibly damaging	Tolerated	Novel
ABCB1	c.G3385A	p.E1129K	1/191	Benign	Damaging	Novel
CYP2C19	c.C629A	p.T210N	1/191	Benign	Damaging	Novel
CYP2C19	c.A818C	p.K273T	1/191	Benign	Damaging	Novel

[1] Annotations were based on NM_003057, NM_153187 for *SLC22A1*, NM_006446 for *SLCO1B1*, NM_000927 for *ABCB1* and NM_000769 for *CYP2C19*.

[2] Novelty: not previously reported in dbSNP or 1000 GP.

[3] Different allele was reported previously with rs ID, rs376060151 (G/A).

[4] Different allele was reported in ABCMdb,[41] in several publications regarding p.G251V.

4. *In vitro* functional assay showed function-altering effects of genetic variants in four ADME genes (*SLC22A1, SLCO1B1, ABCB1 and CYP2C19*)

Next, we evaluated functional consequences of unreported rare and low-frequency variants. We prioritized target genes for *in vitro* assays among Phase I metabolizing enzymes and influx/efflux transporters. Then, we investigated clinical evidences of each gene paired to drugs from Pharmacogenomics Knowledge Base (PharmGKB)[42] and the FDA Table of Pharmacogenomic Biomarkers in Drug Labels. Among those genes, 4 ADME genes were chosen as reference substrates were available. Importantly, CYP2C19 was the second most frequently associated gene with multiple drugs among FDA-informed pharmacogenomic biomarkers. The rest of three genes of influx/efflux transporters were associated with multiple drugs as clinical evidences to varying degrees (Figure 10).

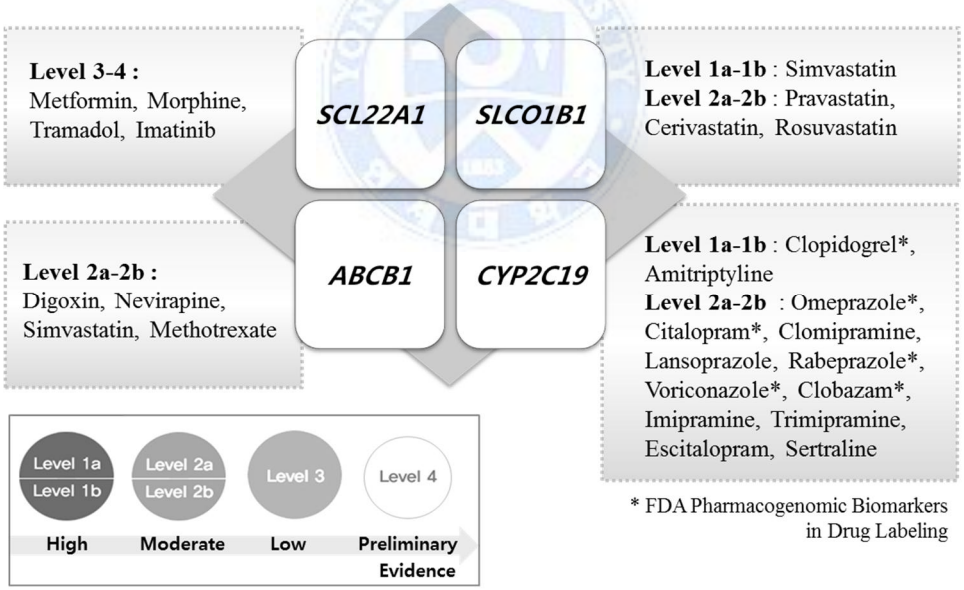

Figure 10. Clinical relevance of 4 ADME genes to drugs. Data was extracted from public databases (PharmGKB and FDA pharmacogenomic biomarkers). The strength of evidence (bottom left) was presented as PharmGKB. Genetic markers of Level 1a represent established markers, as included in dosing guidelines.

Selected rare and low-frequency variants for 4 genes were tested with the reference substrates and most of the variants significantly reduced intrinsic functions of proteins (Figure 11). Two of three variants in SLC22A1/OCT1 consistently reduced the uptake of [14C]-TEA in HEK 293 cells that were transfected with mutated OCT1 plasmids (Figure 11A). Because OCT1(SLC22A1) primarily functions in the hepatic uptake of cationic drugs,[43] function-reducing variants such as p.A33V and p.Y91V may reduce the hepatic excretion and alter the exposure and response to OCT1 substrate drugs. Since these two variants are not directly related to forming substrate binding pockets,[44] reduced function of OCT1 can affect the wide variety of OCT1 substrate drugs. Besides, consequences of two variants in SLCO1B1/OATP1B1 were as predicted by *in silico* predictions. Only p.V84L variant significantly reduced the influx of [3H]-ES depending on its concentration (Figure 11B). As implied from statin-induced myopathy by common genetic variants,[45] the decreased function of OATP1B1 by p.V84L can be possibly associated with ADRs of substrate drugs. Also, two of three variants in ABCB1/MDR1 reduced the efflux of Calcein, resulting in the increased intracellular fluorescence (Figure 11C). Since MDR1 is one of the largely distributed proteins in the body,[46] the diminished function may cause the direct influences not only on absorption but also elimination of MDR1 substrate drugs. Furthermore, two variants of CYP2C19 significantly lowered 4'-hydroxylation function for (S)-Mephenytoin (Figure 11D). Two variants are likely to affect the catalytic efficiency of CYP2C19 for other substrate drugs as well because two sites (Thr210 and Lys273) are not within the specific binding sites of (S)-Mephenytoin.[47]

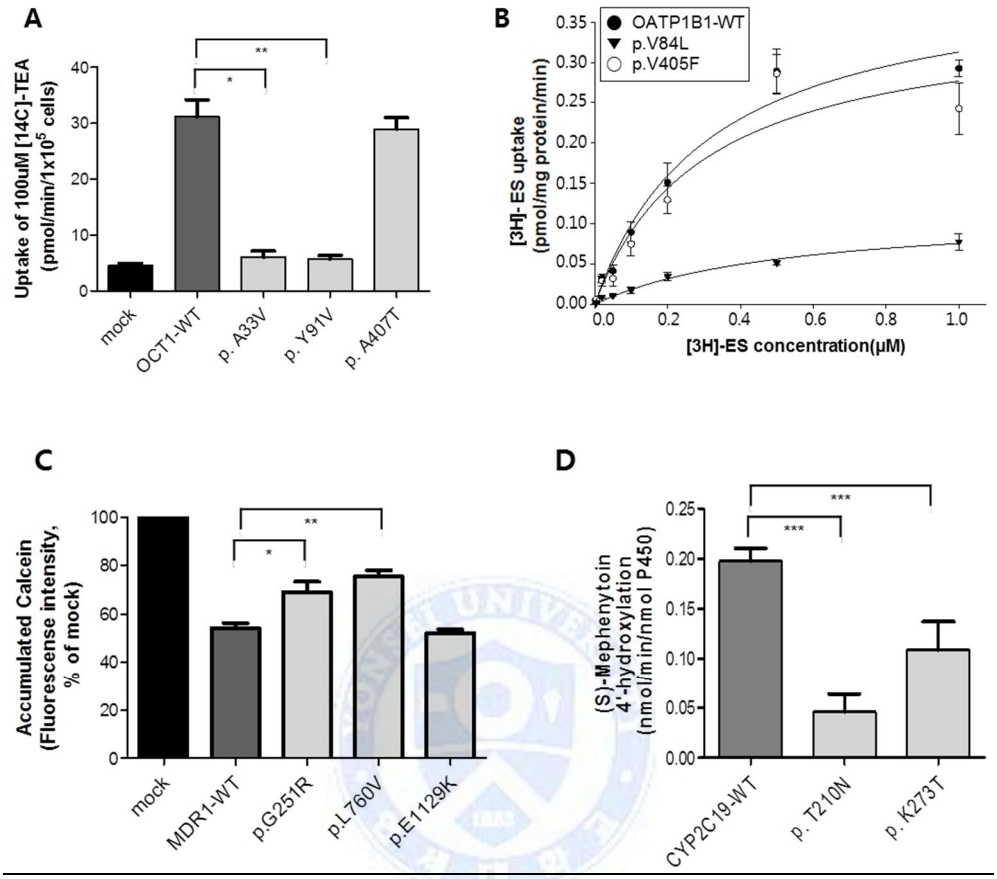

Figure 11. ***In vitro*** **functional evaluation of rare and low-frequency variants in 4 ADME genes.** (A) Uptake of 100μM [14C]-TEA by a wild-type and three mutated OCT1 proteins (p.A33V, p.Y91V, p.A407T) in transiently transfected HEK293 cells. (B) Concentration of [3H]-ES dependency by a wild-type and two mutated OATP1B1 proteins (p.V84L, p.V405F) in transiently transfected HEK293T cells. (C) Accumulated Calcein (%) compared to mock between a wild-type and three mutated MDR1 proteins (p.G251R, p.L760V, p.E1129K). The efflux of MDR1 protein is conversely proportional to accumulated Calcein, which was measured by fluorescence intensity. (D) 4'-Hydroxylation of (S)-Mephenytoin by a wild-type and two mutated CYP2C19 proteins (p.T210N, p.K273T). Results were shown as mean ± SEM. * $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$, WT , Wild-Type.

5. Targeted sequencing identified novel genetic determinants for the PK variability of Tacrolimus

Tacrolimus is an immunosuppressive drug with the large inter-individual variances in PK profiles. Therefore, defining genetic determinants of PK variability may help adjusting proper dosing of Tacrolimus at an earlier stage. As reported previously, Coefficient of Variations (CV) of AUC_{last} values was high in the two cohorts, which were 48.86% and 59.24%, respectively (Figure 12A and 12C). Target capture of 43 healthy volunteers in Phase 1 clinical trial of 1mg Tacrolimus showed that substantial portions of inter-individual variability in AUC_{last} was accounted by CYP3A5*3 (rs776746, c.6986T>C). As expected from previous studies,[48] CYP3A5*3 was the top component that can explain the variations in drug exposure (Figure 12B). The effect of the functionally impaired CYP3A5*3 allele on AUC_{last} was also replicated in an independent cohort of 47 healthy individuals taking 2 capsules of 1mg Tacrolimus (Figure 12D). The CYP3A5*3 allele has been largely investigated, partly because the variants is common in the most of population. Therefore, for the drugs metabolized by CYP3A5, metabolizer types by CYP3A5*3 allele can be categorized into extensive, intermediate and poor metabolizers (EM, IM, and PM).[33] Several variants showed the same trend to explain the variability of AUC_{last} since they are in linkage disequilibrium with rs776746 (Figure 13A-F). Nevertheless, a part of people in PK outliers and extremes were not explained by CYP3A5*3 solely. In prior screening of genes with known associations such as three canonical PK genes and other genes, only p.A503V (rs1057868, c.C1508T) of P450 oxidoreductase (POR) gene was confirmed to reduce the drug exposure specifically for CYP3A5 expressers (Figure 14). This was consistent with the previous reports of POR*28.[49-51] Since POR activity is highly coupled with the oxidative activity by CYP enzymes, the p.A503V variant may facilitate the enzyme activity by activating electron transfers.

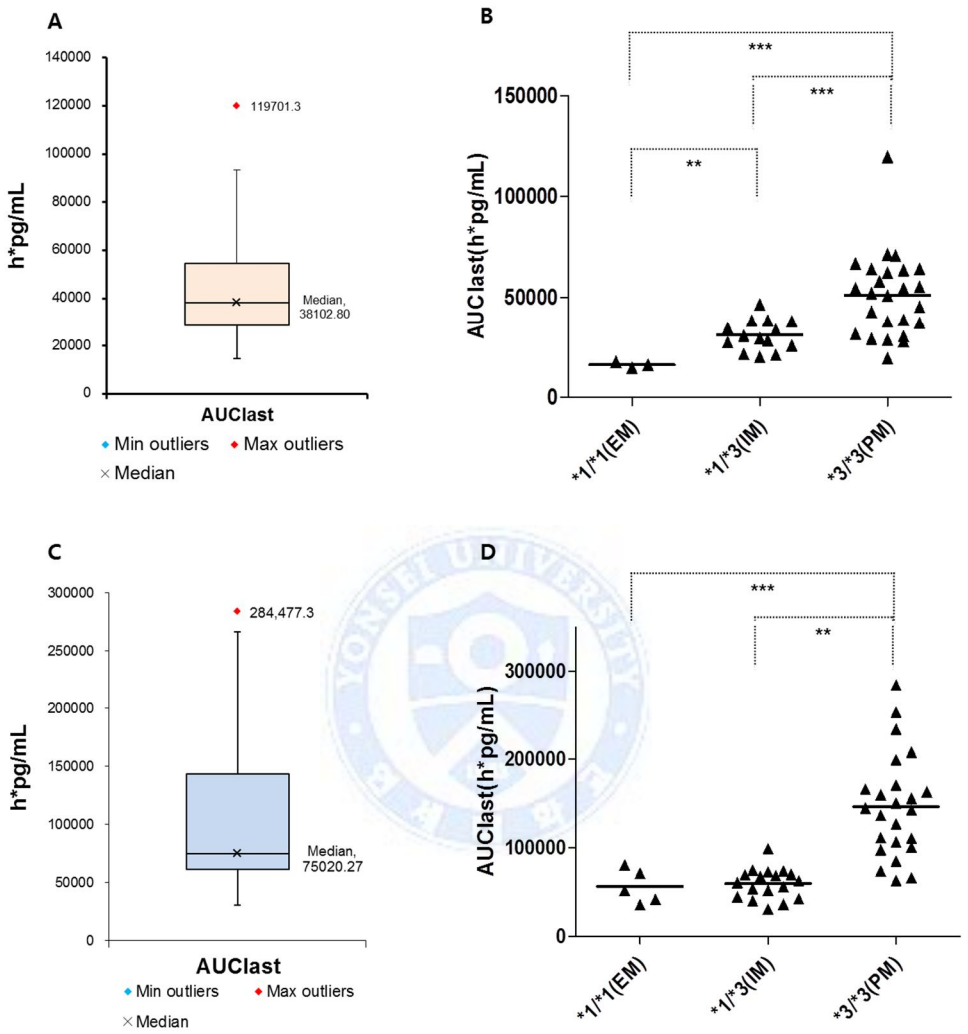

Figure 12. Variable drug exposure (AUC_{last}) explained by metabolizing groups by CYP3A5*3 genotypes. The variability in AUC_{last} was high as 47.86% (A, *n*=43) and 59.24% (C, *n*=47) when measured at CV. The CYP3A5*3 variant accounted for the variability of AUC_{last} with statistical significances in two cohorts (B, *n*=43 and D, *n*=47). * $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$. EM, Extensive Metabolizer; IM, Intermediate Metabolizer; PM, Poor Metabolizer.

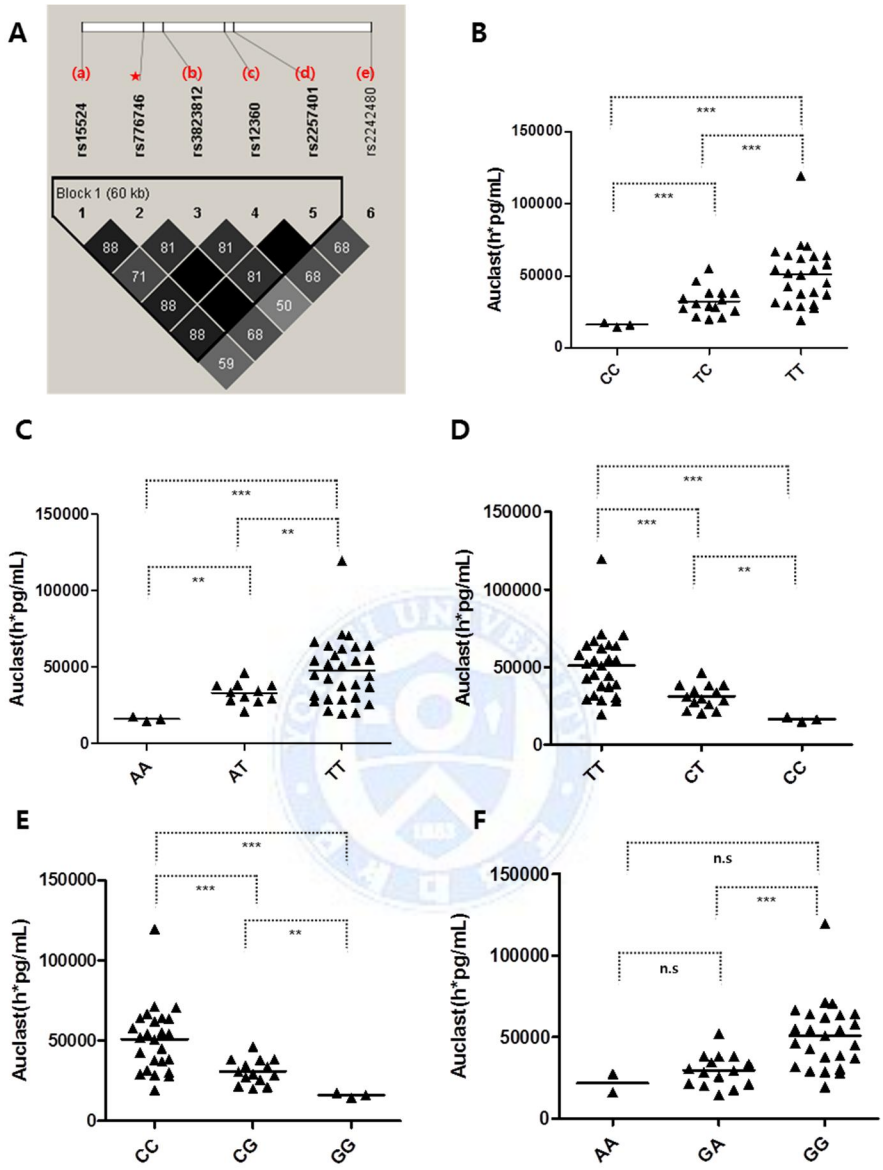

Figure 13. SNPs linked with CYP3A5*3(rs776746) and PK of Tacrolimus.

(A) LD blocks near CYP3A5*3 (rs776746, marked with an asterisk). Variants within ± 500kb of rs776747 were investigated from 43 samples and R'-square values were depicted in each block. These variants showed the same trend as CYP3A5*3 allele did with statistical significances at varying degrees. The plots

showed categorizing of AUC_{last} by rs15524 ((B), CYP3A5, C>T), rs3823812 ((C), CYP3A5, A>T), rs12360 ((D), CYP3A7, G>A), rs2257401 ((E), CYP3A7*2, C>G, p.T409R) and rs2242480 ((F), CYP3A4*1G, A>G). * $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$.

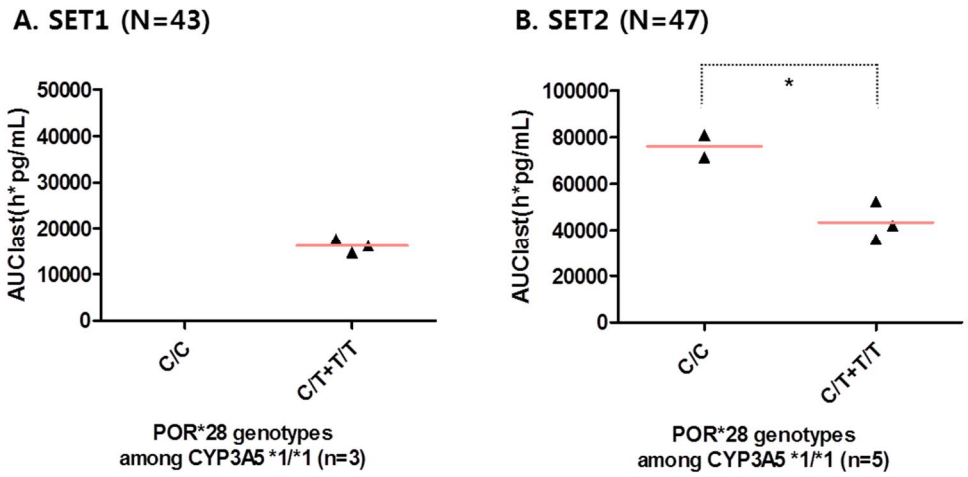

Figure 14. The POR*28 T allele of reducing drug exposure to Tacrolimus among CYP3A5 full expressers. Whereas the POR*28 CC genotype was absent in three EM of 43 samples (A), POR*28 T allele carriers of three EM of 47 samples showed less exposures to Tacrolimus than non-carriers (B). * $P <$ 0.05.

We further investigated the unknown genetic variants for associations with fluctuating PK profiles. There was no significant association with any single variant for both sets. For rare variants of MAF ≤ 0.01, we performed the gene-based burden test by comparing means between carrier and non-carrier of rare variants within pooled 90 samples (Table 4). The test of nonsynonymous variants revealed that rare variants in UGT1A4 had the significantly different AUC_{last} between carriers and non-carriers (p = 0.0003 at MAF cutoff = 0.01). These rare variants in UGT1A4 also influenced drug exposure within the PM group with the robust trend toward significance (p = 0.054 at MAF cutoff = 0.01). In the detailed examination of these variants in each set of 43 and 47, we discovered a number of rare variants (MAF ≤ 0.01) in phenotypic extremes of 3 metabolizer groups by CYP3A5*3 (Figure 15). We discovered that upper phenotypic extremes of PM and IM in 43 samples shared one stop-gain variant (p.Q98X). Notably, the p.Q98X carrier of CYP3A5*3 IM was compound heterozygous with two additional frame-shift and missense variants. The p.L132P variant was found in a lower extreme of PM in 43 samples and expected to decrease drug exposure by increasing phase II metabolism of Tacrolimus. One additional p.R11W variant was found in one of EM but no significant effect on PK variability was detected. Though p.R11W (UGT1A4*4) showed the increased activity for glucuronidation of Tacrolimus,[28] the effect within EM by CYP3A5*3 was insignificant. This suggested that UGT1A4 functions distinctively in the context of functionally disrupted CYP3A5. In a set of 47 samples, two rare variants (p.E50D and p.A58V) were discovered in lower extremes of PM and IM, respectively. These variants appeared in less than 0.1% of population, according to 1000GP as listed (Table5).

Taken together, the PK variance can be explained more by POR*28 and rare variants of UGT1A4 (R^2 = 37.7% and 47.7 %) than solely by CYP3A5*3 (R^2 = 31.8% and 41.7 %, respectively for 43 and 47 samples). To conclude, we could increase approximately 6% of explanatory information for the PK variability of Tacrolimus.

Table 4. Gene-based burden test for rare variants of MAF 1%

Gene	Group	T-test statistics	Mean of dose-adjusted AUC_{last}	
		P-value	Variants Carrier	Variants Non-carrier
UGT1A4	All	0.0003	24432.1 (5)	48434.5 (85)
	*3/*3 PM	0.054	25490.4 (2)	63032.1 (48)
	*1/*3 IM	0.518	26775.9 (2)	30646.9 (30)
	*1/*1 EM	NA	17627.6 (1)	24569.2 (7)
CYP2B6	All	6.40×10^{-5}	35327.5 (2)	47368.6 (88)
	*3/*3 PM	NA	NA (0)	61530.5 (50)
	*1/*3 IM	NA	35011.4 (1)	30256.3 (31)
	*1/*1 EM	NA	35643.6 (1)	21995.5 (7)
SCN5A	All	0.029	69741.0 (6)	45483.9 (84)
	*3/*3 PM	0.292	73785.7 (5)	60168.8 (45)
	*1/*3 IM	NA	49517.4 (1)	29788.4 (31)
	*1/*1 EM	NA	NA (0)	23701.5 (8)
POR	All	0.031	35563.8 (6)	47925.1 (84)
	*3/*3 PM	NA	55368.6 (1)	61656.2 (49)
	*1/*3 IM	0.501	31602.8 (5)	30183.1 (27)
	*1/*1 EM	NA	NA (0)	23701.5 (8)
BRCA1	All	4.67×10^{-14}	22011.7 (2)	47671.2 (88)
	*3/*3 PM	NA	NA (0)	61530.5 (50)
	*1/*3 IM	0.129	22011.7 (2)	30964.5 (30)
	*1/*1 EM	NA	NA (0)	23701.5 (8)
EGFR	All	0.202	63697.4 (4)	46329.1 (86)
	*3/*3 PM	0.207	85442.2 (2)	60534.1 (48)
	*1/*3 IM	0.033	41952.6 (2)	29635.1 (30)
	*1/*1 EM	NA	NA (0)	23701.5 (8)

Only genes that showed statistical significance (P-value < 0.05) in any group were listed. The number in parentheses represents the number of samples in carriers or non-carriers of variants of MAF 1%. PM, Poor Metabolizer; IM, Intermediate Metabolizer; EM, Extensive Metabolizer, NA; Not Available because less than one sample was detected

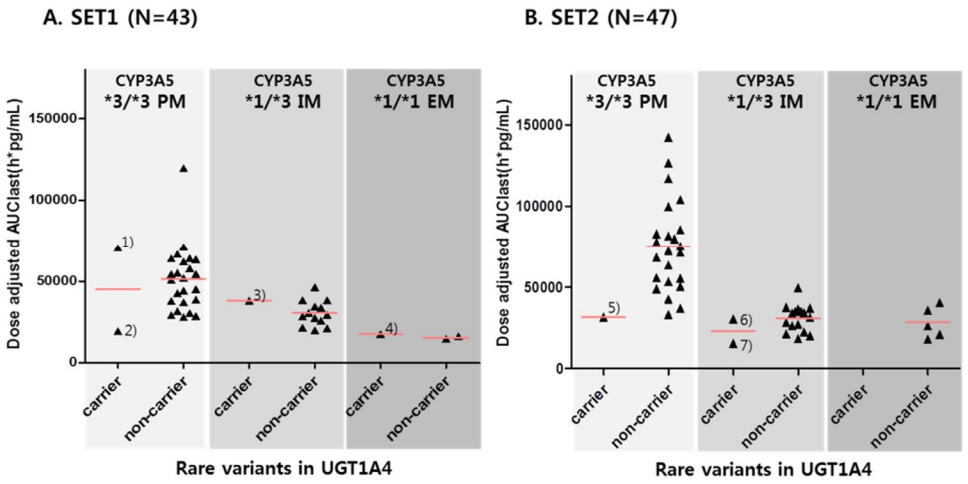

Figure 15. Rare variants in UGT1A4 carried by phenotypic extremes. Rare variants with bi-directional effects in UGT1A4 functions showed different drug exposure in phenotypic extremes of both sets. (A) The p.Q98X[1)] and compound heterozygote[3)] of p.Q98X and p.A58fs were discovered in two upper extremes of PM and IM. The p.L132P[2)] was identified in the lower extreme of PM. One additional p.R11W[4)] was found in one of EM but no significant effect was shown. (B) Three rare variants (p.E50D[5)], p.E454Q[6)] and p.A58V[7)]) were found in lower extremes of PM and IM, respectively.

Table 5. The detailed information of rare variants in UGT1A4

	Variant	Protein	rs ID	AF [1]	Carrier (Set [2])
1)	c.C292T	p.Q98X	rs201323245	0.0004	Upper extreme of PM (43)
2)	c.T395C	p.L132P	rs72551337	0.0002	Lower extreme of PM (43)
3)	c.[C292T]+ [74delG ; A325G]	p.[A58fs] + [Q98X]	NA	NA [3]	Upper extreme of IM (43)
4)	c.C31T	p.R11W	rs3892221	0.0140	One of EM (43)
5)	c.G150C	p.E50D	rs45510694	0.001	Lower extreme of PM (47)
6)	c.G1360C	p.E454Q	NA	NA	One of IM (47)
7)	c.C173T	p.A58V	rs141408391	0.002	Lower extreme of IM (47)

The numbers in the first column are matched with ones in figure 15. AF, Allele Frequency; NA, Not Available.

[1] Allele frequency screened from 1000 GP.

[2] The number in parentheses represents the number of samples for the cohorts, of which corresponding carrier are included in.

[3] The c. A325G (rs539093785) appeared in 0.02% of 1000 GP. However, the compound heterozygote of these three variants has not been reported yet.

IV. DISCUSSION

In the present study, we developed the optimized probe set based on MIP, which can be readily applicable for pharmacogenomics analysis. Through modifications from previously reported MIP methods, this platform featured less costly and laborious preparation protocol. The genotyping performance was comparable to the existing platforms as measured by PPV and sensitivity. Nevertheless, there was a part of genes that were not perfectly captured by this panel. We speculated that the sequence similarity of ADME genes might have impeded the capture performance of a part of probes. We expected that further serial rebalancing could have improved the coverage significantly as we achieved in the first rebalancing. One drawback of microarray-based MIP synthesis is that it is impossible to sort poorly performing MIPs out from all probes. Therefore, we mixed the original probes with the rebalanced probes without eradicating poorly performing MIPs. We expected that a panel of selected probes with the high performance would improve the coverage a lot higher in the future.

Target capture and NGS of DNA sequences from 191 Koreans revealed unreported variants, especially ones in low frequency. Although 191 Koreans were not sufficient to discover very rare variants, the trend of higher ratio of nonsynonymous variants at low frequency could be observed. Moreover, *in silico* prediction programs expected that genetic variants in lower frequency were enriched for those with damaging effects. It is not to say all variants in low frequency were deleterious, but *in vitro* evaluation of 4 ADME genes implicates genetic variants at low frequency, which might have been missed out in past pharmacogenetic studies, may account for variable drug exposure and response. Still, it is difficult to decide how in vitro results can be reflected in clinical practice. Evaluating the clinical meaning of newly discovered rare variants remains as an issue to adopt for decision making in clinical practice.

Application of the optimized MIP set to the participants in phase 1 clinical

trials proved the utility of this platform. Aside from CYP3A5*3, an established index variant for the variability, POR*28 and rare variants in UGT1A4 were revealed to be the account for the part of variable drug exposure to Tacrolimus. The drug exposure by POR*28 was significantly decreased only in the context of fully functional CYP3A5. On the other hand, rare variants in UGT1A4 showed bi-directional effects on drug exposure and significant influence in the context of functionally impaired CYP3A5. Interestingly, while p.L48V (UGT1A4*3), a commonly discovered variant in both sets, showed no significant functional effects on drug exposure, only variants of low frequency had impacts on the extent of drug exposure. To the best of our knowledge, this is the first time to report the association of rare genetic variations of UGT1A4 with the PK of Tacrolimus. Taken together, we expected to increase approximately 6% of explanatory information by adopting POR*28 and pooled rare variants of UGT1A4 as genetic determinants of a model for the PK variability of Tacrolimus.

Though UGT1A4 was pointed as a major contributor of Tacrolimus glucuronidation,[28] it is still unclear that how glucuronidation by UGT1A4 interplays with oxidative metabolisms by CYP 3A enzymes. In addition, the conjugated drugs to glucuronide could be recycled via enterohepatic circulation, affecting the drug exposure up to the total clearance of drugs. To further explore the extent of contribution of glucuronidation by UGT1A4, the difference in enzyme kinetics by mutated proteins should be evaluated *in vitro*. Also, the sample size was insufficient to discover a great diversity of rare variants in current cohorts, which resulted in discovering only suggestive trend toward significance in association tests. Therefore, further validation of our findings in larger cohorts of healthy volunteers and patients is needed.

V. CONCLUSION

The present study showed that the optimized probe set based on MIP is a readily applicable platform to capture common markers and unrevealed rare variants in ADME and PD genes for pharmacogenomic analysis. Target capture and NGS of DNA sequences from 191 Koreans showed the unreported low-frequency variants were enriched for nonsynonymous variants especially with potentially damaging effects. Moreover, relatively rare variants might have been underestimated in previous pharmacogenetic studies, in spite of functional impacts, as shown by *in vitro* functional validation in 4 ADME genes. When applying the panel to healthy volunteers in phase 1 clinical trials, beyond known markers, novel genetic determinants were identified as contributing factors to variable extents of exposure to Tacrolimus. One of notable findings was the contribution of rare variants in UGT1A4 for PK pathway of Tacrolimus. Taken together, these results implicate that not only common but also rare variants are worth being considered for pharmacogenomic analysis. We also present a proof of concept that NGS-based screening of pharmacogenes at earlier phase of clinical trials can help understanding novel metabolic pathways of drugs by taking unknown rare and low-frequency genetic variations into account.

REFERENCES

1. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods 2010;7:111-8.
2. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res 2013;23:843-54.
3. Yoon JK, Ahn J, Kim HS, Han SM, Jang H, Lee MG, et al. microDuMIP: target-enrichment technique for microarray-based duplex molecular inversion probes. Nucleic Acids Res 2015;43:e28.
4. Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat Genet 1998;19:225-32.
5. Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, Brundage T, et al. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. Genome Res 2005;15:269-75.
6. Dumaual C, Miao X, Daly TM, Bruckner C, Njau R, Fu DJ, et al. Comprehensive assessment of metabolic enzyme and transporter genes using the Affymetrix Targeted Genotyping System. Pharmacogenomics 2007;8:293-305.
7. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 2012;338:1619-22.
8. Shen P, Wang W, Chi AK, Fan Y, Davis RW, Scharfe C. Multiplex target capture with double-stranded DNA probes. Genome Med 2013;5:50.
9. Carlson KD, Sudmant PH, Press MO, Eichler EE, Shendure J, Queitsch C. MIPSTR: a method for multiplex genotyping of germline and somatic STR

variation across many individuals. Genome Res 2015;25:750-61.

10. Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, Cook J, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet 2013;45:825-30.
11. Daly AK. Genome-wide association studies in pharmacogenomics. Nat Rev Genet 2010;11:241-6.
12. Motsinger-Reif AA, Jorgenson E, Relling MV, Kroetz DL, Weinshilboum R, Cox NJ, et al. Genome-wide association studies in pharmacogenomics: successes and lessons. Pharmacogenet Genomics 2013;23:383-94.
13. Altman RB, Kroemer HK, McCarty CA, Ratain MJ, Roden D. Pharmacogenomics: will the promise be fulfilled? Nat Rev Genet 2011;12:69-73.
14. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 2012;337:100-4.
15. Ramsey LB, Bruun GH, Yang W, Trevino LR, Vattathil S, Scheet P, et al. Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. Genome Res 2012;22:1-8.
16. Hewett M, Oliver DE, Rubin DL, Easton KL, Stuart JM, Altman RB, et al. PharmGKB: the Pharmacogenetics Knowledge Base. Nucleic Acids Res 2002;30:163-5.
17. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clin Pharmacol Ther 2011;89:464-7.
18. Tutton R. Pharmacogenomic biomarkers in drug labels: what do they tell us? Pharmacogenomics 2014;15:297-304.
19. Vivot A, Boutron I, Ravaud P, Porcher R. Guidance for pharmacogenomic biomarker testing in labels of FDA-approved drugs. Genet Med 2015;17:733-8.

20. Ventola CL. Pharmacogenomics in clinical practice: reality and expectations. P t 2011;36:412-50.
21. Van Driest SL, Shi Y, Bowton EA, Schildcrout JS, Peterson JF, Pulley J, et al. Clinically actionable genotypes among 10,000 patients with preemptive pharmacogenomic testing. Clin Pharmacol Ther 2014;95:423-31.
22. Ratain MJ, Johnson JA. Meaningful use of pharmacogenetics. Clin Pharmacol Ther 2014;96:650-2.
23. Janssens AC, Deverka PA. Useless until proven effective: the clinical utility of preemptive pharmacogenetic testing. Clin Pharmacol Ther 2014;96:652-4.
24. Rodriguez-Peralvarez M, Germani G, Darius T, Lerut J, Tsochatzis E, Burroughs AK. Tacrolimus trough levels, rejection and renal impairment in liver transplantation: a systematic review and meta-analysis. Am J Transplant 2012;12:2797-814.
25. Staatz C, Taylor P, Tett S. Low tacrolimus concentrations and increased risk of early acute rejection in adult renal transplantation. Nephrol Dial Transplant 2001;16:1905-9.
26. Dai Y, Hebert MF, Isoherranen N, Davis CL, Marsh C, Shen DD, et al. Effect of CYP3A5 polymorphism on tacrolimus metabolic clearance in vitro. Drug Metab Dispos 2006;34:836-47.
27. Firdaous I, Verbeeck RK, Hassoun A, Langrehr JM, Wallemacq PE. Excretion of tacrolimus glucuronides in human bile. Eur J Drug Metab Pharmacokinet 1997;22:217-21.
28. Laverdiere I, Caron P, Harvey M, Levesque E, Guillemette C. In vitro investigation of human UDP-glucuronosyltransferase isoforms responsible for tacrolimus glucuronidation: predominant contribution of UGT1A4. Drug Metab Dispos 2011;39:1127-30.
29. Strassburg CP, Barut A, Obermayer-Straub P, Li Q, Nguyen N, Tukey RH, et al. Identification of cyclosporine A and tacrolimus glucuronidation in

human liver and the gastrointestinal tract by a differentially expressed UDP-glucuronosyltransferase: UGT2B7. J Hepatol 2001;34:865-72.

30. Cardenas ME, Hemenway C, Muir RS, Ye R, Fiorentino D, Heitman J. Immunophilins interact with calcineurin in the absence of exogenous immunosuppressive ligands. Embo j 1994;13:5944-57.

31. Shaw KT, Ho AM, Raghavan A, Kim J, Jain J, Park J, et al. Immunosuppressive drugs prevent a rapid dephosphorylation of transcription factor NFAT1 in stimulated immune cells. Proc Natl Acad Sci U S A 1995;92:11205-9.

32. Clipstone NA, Crabtree GR. Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. Nature 1992;357:695-7.

33. Birdwell KA, Decker B, Barbarino JM, Peterson JF, Stein CM, Sadee W, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP3A5 Genotype and Tacrolimus Dosing. Clin Pharmacol Ther 2015;98:19-24.

34. Williams R, Peisajovich SG, Miller OJ, Magdassi S, Tawfik DS, Griffiths AD. Amplification of complex gene libraries by emulsion PCR. Nat Methods 2006;3:545-50.

35. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. 2014;32:246-51.

36. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.

37. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.

38. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat

Methods 2010;7:248-9.

39. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

40. Yoon JK, Ahn J, Kim HS, Han SM, Jang H, Lee MG, et al. microDuMIP: target-enrichment technique for microarray-based duplex molecular inversion probes. Nucleic Acids Res 2014.

41. Gyimesi G, Borsodi D, Saranko H, Tordai H, Sarkadi B, Hegedus T. ABCMdb: a database for the comparative analysis of protein mutations in ABC transporters, and a potential framework for a general application. Hum Mutat 2012;33:1547-56.

42. Thorn CF, Klein TE, Altman RB. PharmGKB: the pharmacogenetics and pharmacogenomics knowledge base. Methods Mol Biol 2005;311:179-91.

43. Koepsell H. Polyspecific organic cation transporters: their functions and interactions with drugs. Trends Pharmacol Sci 2004;25:375-81.

44. Koepsell H, Schmitt BM, Gorboulev V. Organic cation transporters. Rev Physiol Biochem Pharmacol 2003;150:36-90.

45. Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, et al. SLCO1B1 variants and statin-induced myopathy--a genomewide study. N Engl J Med 2008;359:789-99.

46. Sakaeda T, Nakamura T, Okumura K. Pharmacogenetics of MDR1 and its impact on the pharmacokinetics and pharmacodynamics of drugs. Pharmacogenomics 2003;4:397-410.

47. 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 2004;21:2270-8.

48. Hesselink DA, van Schaik RH, van der Heiden IP, van der Werf M, Gregoor PJ, Lindemans J, et al. Genetic polymorphisms of the CYP3A4, CYP3A5, and MDR-1 genes and pharmacokinetics of the calcineurin inhibitors

cyclosporine and tacrolimus. Clin Pharmacol Ther 2003;74:245-54.

49. de Jonge H, Metalidis C, Naesens M, Lambrechts D, Kuypers DR. The P450 oxidoreductase *28 SNP is associated with low initial tacrolimus exposure and increased dose requirements in CYP3A5-expressing renal recipients. Pharmacogenomics 2011;12:1281-91.

50. Zhang JJ, Zhang H, Ding XL, Ma S, Miao LY. Effect of the P450 oxidoreductase 28 polymorphism on the pharmacokinetics of tacrolimus in Chinese healthy male volunteers. Eur J Clin Pharmacol 2013;69:807-12.

51. Elens L, Hesselink DA, Bouamar R, Budde K, de Fijter JW, De Meyer M, et al. Impact of POR*28 on the pharmacokinetics of tacrolimus and cyclosporine A in renal transplant patients. Ther Drug Monit 2014;36:71-9.

ABTRACT (IN KOREAN)

분자 도치 프로브 (MIP) 기반 다중 타깃 포획을 이용한 약물유전체 분석

<지도교수 이 민 구>

연세대학교 대학원 의과학과

한 수 민

분자 도치 프로브 (MIP)에 기반한 타깃 포획 방법을 응용하여, 약물유전체 분석을 위한 타깃 시퀀싱에 적합화한 프로브 세트를 개발하였다. 이 플랫폼의 프로브들은 80개의 약물의 흡수, 분포, 대사, 배설 및 반응에 중요한 유전자와 현재까지 잘 알려진 유전학적 바이오 마커를 포획하도록 디자인되었다. 반복적인 재조정을 통하여 적합화된 이 플랫폼을 이용하여 96%의 타깃 유전자와 바이오 마커에 대한 유전형을 99.8%의 정확도로 얻을 수 있었다. 이 플랫폼은 microarray에서 합성한 후 증폭한 이중 서열의 분자 도치 프로브 방법을 응용한 것으로, 프로브를 위한 준비가 간편하고 가격이 저렴한 장점이 있다. 이 플랫폼을 191명의 건강한 한국인에 적용하여 목표 유전자와 마커들을 시퀀싱한 결과, 약물의 체내

동태와 관련된 유전자에서, 보고된 바 없는 드문 빈도로 나타나는 변이들을 발견할 수 있었다. 기능적 변화를 예측하는 컴퓨터 프로그램을 이용하여 발견된 변이에 의한 효과를 예측하였을 때, 비교적 드문 빈도로 나타나는 유전자 변이들이 기능적 변화를 더 크게 일으킬 것으로 예측되었다. 이 중 4개의 약물 수송체 및 대사 효소 유전자에서 나타나는 드문 빈도의 유전자 변이들을 골라, 체외(*in vitro*) 실험에서 실제 변이가 있는 단백질의 기능이 변화되는지 확인하였다. 대부분의 변이들이 단백질이 표준 기질을 수송하거나 대사시키는 기능을 저해하는 것을 확인할 수 있었다. 이 결과를 통하여 기존의 약물유전학 연구에서 중요성을 두지 않았던 드문 변이 또한 약물의 반응 및 노출 정도에 영향을 미칠 가능성이 있다는 것을 알 수 있었다. 개인간 약물의 동태 및 반응 차이가 크다고 알려진 약물의 유전적 소인을 분석하기 위하여, 면역억제제인 Tacrolimus를 투여받은 90명의 1단계 임상 시험의 참여자에 이 플랫폼을 적용하였다. 알려진 바와 같이 CYP3A5 유전자에 *3로 명명된 변이에 의해 약물 동태가 통계적으로 유의하게 차이 나는 것을 확인하였다. 이 외에도, POR*28과 UGT1A4의 드문 변이들이 개인간 약물 동태의 차이를 설명할 수 있음을 확인하였다. POR*28은 특히 CYP3A5의 발현이 정상적인 그룹에서 약물에 대한 노출 정도를 줄이는 것을 확인하였다. 반면, UGT1A4의 드문 변이들은 특히 CYP3A5의 발현이 적은 그룹 내의 약물 동태의 차이가 크게 나타나는 피험자에게서 발견되었다. UGT1A4는 Tacrolimus의 glucuronidation에 주요한 효소로 알려져 있기 때문에, 변이에 의한 효소 기능의 차이가 체내 약물 동태의 차이를 유발하였다고 생각할 수 있다. 현재까지 알려진 바에 의하면, UGT1A4의 변이가 Tacrolimus의 약물 동태 차이와 연관이 있다는 보고는 없었다. 결론적으로 POR*28과 UGT1A4의 드문 변이들을 고려하면 CYP3A5*3만 고려하였을 때보다 Tacrolimus의 약물 동태의 차이를

약 6% 더 설명할 수 있었다. 이상의 결과를 통하여 약물유전학 연구와 테스트에서 인구 집단에서 흔하게 나타나는 변이뿐만 아니라 드물게 나타나는 변이도 고려될 필요가 있음을 보였다. 또한, 임상시험의 초기 단계에서 이와 같은 약물유전학 연구를 가능하게 하는 플랫폼을 적용함으로써, 임상시험 중인 약물의 새로운 대사 경로나 작용 기전을 이해하는 것을 도울 수 있을 수 있음을 보였다. 덧붙여, 본 연구에서 확립한 플랫폼이 약물유전학적 연구에 바로 적용할 수 있는 효율적인 도구이며, 약물 반응 및 동태의 개인차를 설명하는 데에도 활용이 가능함을 보였다.

핵심되는 말: 약물유전학, 타깃시퀀싱, 분자 도치 프로브 (MIP), 희귀 변이(rare variants), 약물동태에 관련된 유전자들(ADME genes)

PUBLICATION LIST

1. Ahn JW, Kim HS, Yoon JK, Jang H, **Han SM**, Eun S, et al. Identification of somatic mutations in EGFR/KRAS/ALK-negative lung adenocarcinoma in never-smokers. Genome Med 2014;6:18.

2. Yoon JK, Ahn J, Kim HS, **Han SM**, Jang H, Lee MG, et al. microDuMIP: target-enrichment technique for microarray-based duplex molecular inversion probes. Nucleic Acids Res 2015;43:e28.

3. **Han SM**, Hwang B, Park TG, Kim DI, Rhee MY, Lee BK, et al. Genetic testing of Korean familial hypercholesterolemia using whole-exome sequencing. PLoS One 2015;10:e0126706.

4. Kwon M*, **Han SM***, Kim DI, Rhee MY, Lee BK, Ahn YK, et al. Evaluation of polygenic cause in Korean patients with familial hypercholesterolemia - A study supported by Korean Society of Lipidology and Atherosclerosis. Atherosclerosis 2015;242:8-12. (* equal contribution)

5. Shin DG*, **Han SM***, Kim DI, Rhee MY, Lee BK, Ahn YK, et al. Clinical features of familial hypercholesterolemia in Korea: Predictors of pathogenic mutations and coronary artery disease - A study supported by the Korean Society of Lipidology and Atherosclerosis. Atherosclerosis 2015;243:53-8. (* equal contribution)