

Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics

María Santos, MSc¹, Mikko Niemi, PhD², Masahiro Hiratsuka, PhD³,
Masaki Kumondai, BSc³, Magnus Ingelman-Sundberg, PhD⁴, Volker M. Lauschke, PhD⁴ and
Cristina Rodríguez-Antona, PhD^{1,5}

Purpose: Variability in pharmacokinetics and drug response is shaped by single-nucleotide variants (SNVs) as well as copy-number variants (CNVs) in genes with importance for drug absorption, distribution, metabolism, and excretion (ADME). While SNVs have been extensively studied, a systematic assessment of the CNV landscape in ADME genes is lacking.

Methods: We integrated data from 2,504 whole genomes from the 1000 Genomes Project and 59,898 exomes from the Exome Aggregation Consortium to identify CNVs in 208 relevant pharmacogenes.

Results: We describe novel exonic deletions and duplications in 201 (97%) of the pharmacogenes analyzed. The deletions are population-specific and frequencies range from singletons up to 1%, accounting for > 5% of all loss-of-function alleles in up to 42%

of the genes studied. We experimentally confirmed novel deletions in *CYP2C19*, *CYP4F2*, and *SLCO1B3* by Sanger sequencing and validated their allelic frequencies in selected populations.

Conclusion: CNVs are an additional source of pharmacogenetic variability with important implications for drug response and personalized therapy. This, together with the important contribution of rare alleles to the variability of pharmacogenes, emphasizes the necessity of comprehensive next-generation sequencing–based genotype identification for an accurate prediction of the genetic variability of drug pharmacokinetics.

Genet Med advance online publication 26 October 2017

Key Words: CNVs; drug pharmacokinetics; novel gene deletions; personalized medicine; pharmacogenes

INTRODUCTION

Interindividual variability in drug responses caused by a combination of drug–drug interactions, as well as by physiopathologic, environmental, genetic, and epigenetic factors, constitutes a major challenge for clinical practice.¹ In particular, genetic variability in the genes that encode proteins involved in drug absorption, distribution, metabolism, and excretion (ADME) has been shown to impact on drug pharmacokinetics, efficacy, and safety, highlighting the decisive role of genetic variation for treatment success.^{2,3} Based on such genetic variability, the US Food and Drug Administration and the European Medicines Agency provide guidance in the drug Summaries of Product Characteristics to improve clinical pharmacotherapy.

While common single-nucleotide variants (SNVs) have been extensively studied, the vast extent of rare SNVs and small insertions–deletions (indels) has only recently received attention, driven in part by technological advancements.^{4,5} Furthermore, there is growing evidence that copy-number variants (CNVs), defined as duplications or deletions of DNA segments ranging from ~1 kb to 3 Mb, contribute

substantially to phenotypic diversity and disease.^{6,7} Structural variants in which the whole or parts of the open reading frame of a gene is deleted commonly abrogate gene function, while whole-gene duplications can increase gene dosage and functionality. Functionally relevant CNVs have been described in the ADME genes *CYP2A6*, *CYP2D6*, *GSTM1*, *GSTT1*, and *SULT1A1*, with frequencies that substantially differ across human populations and that contribute to interethnic pharmacokinetic differences.⁸ Particularly for *CYP2D6*, not only deletions but also gene duplications and higher-level amplifications have been described that majorly impact human drug response and whose interethnic differences in frequencies are still not fully characterized.⁹

In recent years, next-generation sequencing (NGS) techniques have provided with powerful approaches for CNV detection, with whole-exome sequencing being able to detect CNVs in exonic sequences with a resolution similar to medium-resolution genomic microarrays.¹⁰ The Exome Aggregation Consortium (ExAC) recently released CNV data derived from whole-exome sequencing of 59,898 individuals distributed across six major human populations.¹¹ However,

¹Hereditary Endocrine Cancer Group, Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain; ²Department of Clinical Pharmacology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; ³Laboratory of Pharmacotherapy of Life-Style Related Diseases, Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan; ⁴Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden; ⁵ISCIII Center for Biomedical Research on Rare Diseases (CIBERER), Madrid, Spain. Correspondence: Volker M. Lauschke or Cristina Rodríguez-Antona (volker.lauschke@ki.se or crodriguez@cnio.es)

The last two authors have co-senior authorship.

Submitted 21 June 2017; accepted 7 August 2017; advance online publication 26 October 2017. doi:10.1038/gim.2017.156

no systematic analysis of CNVs in ADME genes has been presented so far, and the impact of these variants on drug response remains unknown.

By leveraging these novel NGS data sets, we here provide the first panorama of CNVs across 208 important ADME genes and estimate the contribution of these newly described structural variants to the variability in drug response. Furthermore, we experimentally identified the genomic breakpoints of three novel deletions in *CYP2C19*, *CYP4F2*, and *SLCO1B3*, and assessed the allele frequencies of these deletions of potential clinical relevance in 1,080 Spanish, 465 Finnish, and 590 Japanese individuals.

MATERIALS AND METHODS

Data collection

Bioinformatic analyses of CNVs were performed on 208 ADME genes with importance for drug pharmacokinetics (Novel deletions and duplications in 208 pharmacogenes across six populations; **Supplementary Table S1** online). CNV allele counts were analyzed in six major populations (non-Finnish Europeans, Finns, Africans, South Asians, East Asians, and admixed Americans) by integrating CNV data from ExAC containing exome sequences from 59,898 unrelated individuals¹¹ with whole-genome sequencing data from 2,504 individuals provided by the 1000 Genomes Project phase 3.¹² The minor allele frequency (MAF) of rare CNVs was extracted from ExAC (MAF \leq 0.5%), while data from the 1000 Genomes Project was used to collect data for common CNVs.

CNVs analyses

Novel CNVs were defined as those not previously reported in the literature. To calculate CNV MAF from carrier counts, Hardy–Weinberg equilibrium was assumed. For functional predictions of CNV effects, we assumed that the deletion of one or more exons of a gene results in a nonfunctional protein product. To estimate the contribution of deletions to the total number of loss-of-function (LOF) alleles, CNV data was related to LOF alleles derived from SNVs and indels obtained from ExAC as previously described.¹³ We used a conservative definition of LOF and only considered those variants that resulted in frameshifts, premature stop-codons, loss of start-codons, or altered canonical splice sites. Furthermore, we included well-described LOF variants in *CYP2C19* (*CYP2C19**2, rs4244285), *CYP2D6* (*CYP2D6**41, rs28371725), and *CYP3A5* (*CYP3A5**3, rs776746) that are either misclassified as nondefective variants or not covered by exome sequencing.

Breakpoint determination

DNA from three individuals (HG00268, HG01485, and NA19010) was acquired from the Coriell Biorepository (Coriell Cell Repositories, Camden, NJ) to identify deletion breakpoints corresponding to *CYP2C19*-esv3624259, *SLCO1B3*-esv3628797, and *CYP4F2*-esv3643780. Primers were designed to bind the flanking region of each deletion (*CYP2C19*-Fw: 5'-ATTAGCAATGTTGCCCGAAG-3'; *CYP2C19*-Rv: 5'-AGAAGAGCAACCCCAAGACA-3'; *SLCO1B3*-

Fw: 5'TCCAAACCCACTTTGTTTCC-3'; *SLCO1B3*-Rv: 5'-TGCTGTGGGTGAATTGAAAG-3'; *CYP4F2*-Fw: 5'-AACCACTCATCCCACCACTC-3'; *CYP4F2*-Rv: 5'-TGACGGCAAGGAAATAAAGC-3') and to amplify through polymerase chain reaction (PCR) the region containing the deletion breakpoints. PCR products were purified using the ExoSAP-IT for PCR Product Clean-Up (Thermo Fisher Scientific, Waltham, MA, USA) and subjected to Sanger sequencing using an ABI PRISM 3700 DNA Analyzer capillary sequencer (Thermo Fisher Scientific). DNA sequencing chromatograms were aligned to the reference human genome (GRCh38) to define the genomic coordinates of the deletions.

Samples and genotyping

Germ-line DNA was collected from unrelated individuals from Spain ($n = 1,080$), Finland ($n = 465$), and Japan ($n = 590$). Individuals were over 18 years old, and the collection of samples was approved by local ethical review committees. CNV genotyping was performed using the Kompetitive Allele Specific PCR (KASP) genotyping assays (LGC Genomics, Hoddesdon, UK) using specific assays designed for the *CYP2C19*, *SLCO1B3*, and *CYP4F2* deletion breakpoints detected by Sanger sequencing. The KASP assay consists of two allele-specific forward primers, one labeled with FAM dye and the other with HEX dye and one common reverse primer. KASP reactions were carried according to the manufacturer's instructions. Briefly, reactions were run in 5 μ l final reaction volume containing 2.5 μ l of KASP 2X reaction mix, 0.07 μ l of assay primers mix, and 15 ng of genomic DNA. The following thermal cycling conditions were used: 94 °C for 15 min, followed by 10 touchdown cycles of 94 °C for 20 s, 61–55 °C for 60 s (dropping 0.6 °C per cycle), and then 26 cycles of 94 °C for 20 s, and 55 °C for 60 s. All assays included positive control samples with known genotypes and negative controls. All deletions identified by genotyping were confirmed by PCR using the breakpoint-specific primers described above.

RESULTS

The landscape of ADME gene CNVs

To provide a comprehensive overview of CNVs within clinically relevant pharmacogenes, we collected data from ExAC and the 1000 Genomes Project and identified deletions and duplications. Of the 208 ADME genes analyzed, 201 harbored novel CNVs (97%) and we identified a total of 5,589 novel CNVs, of which 2,611 (47%) were deletions and 2,978 (53%) were duplications (**Figure 1a**).

Deletions were detected in 175 out of 208 genes (84%; **Figure 1b** and **Supplementary Table S1**). Deletions affecting *CYP2C19* in Finns, *CYP2B6* in Africans, and *CYP4F2* in East Asians were the most frequent with MAFs of 1.1%, 0.9%, and 0.4%, respectively. Most of these newly described deletions were highly population-specific, with interpopulation differences in allele frequencies >10-fold, while others were present in all populations studied at comparable levels, such as those affecting *CYP2F1* and

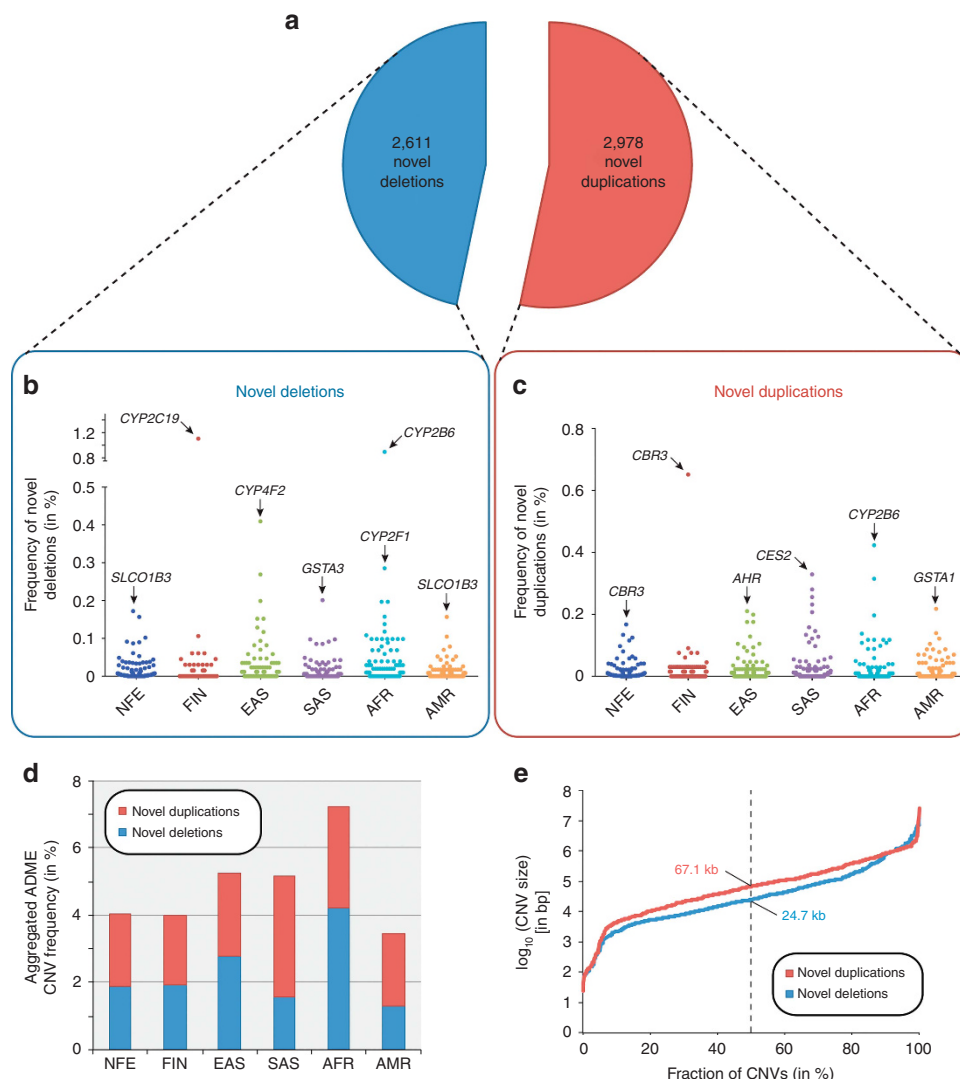


Figure 1 Overview of newly described gene deletions and duplications in 208 pharmacogenes across six human populations. (a) In total we found 5,589 novel copy-number variants (CNVs) in 59,898 individuals of which 2,611 were deletions (blue) and 2,978 duplications (red). (b) Dot plot depicting allele frequencies for novel deletions across non-Finnish Europeans (NFE), Finnish (FIN), East Asians (EAS), South Asians (SAS), Africans (AFR), and admixed Americans (AMR). Deletions of the *CYP2A7* pseudogene as well as previously described deletions are not included in this representation and are depicted in **Supplementary Figure S1** and **Table 1**, respectively. The highest deletion frequencies were observed for *CYP2C19* in Finns and *CYP2B6* in Africans. (c) Allele frequencies of novel duplications are shown. Duplications of *CBR3* and *CYP2B6* in Finnish and African individuals, respectively, were the most common. (d) Aggregated CNV frequencies were highest in Africans (7.2%) and lowest in admixed Americans (3.4%). (e) The size distributions of deletions and duplications are shown. Duplications were slightly larger affecting median genomic intervals of 67.1 kb compared to 24.7 kb for deletions. ADME, absorption, distribution, metabolism, and excretion.

SLCO1B3. Of all CNVs identified, deletions affecting the pseudogene *CYP2A7*, which may act as a micro RNA decoy for miR-126 and affect *CYP2A6* levels,¹⁴ were most abundant with frequencies between 1.7% and 11.5% in the different populations (**Supplementary Figure S1**). Furthermore, we refined the population-specific data of previously well-characterized deletions in *GSTM1*, *GSTT1*, *UGT2B17*, *UGT2B28*, *CYP2D6*, *CYP2A6*, *SULT1A1*, and *CYP2B6* using the 1000 Genomes Project data (**Table 1**). Overall, the aggregated frequency of the novel ADME gene deletions ranged from 4.2% in Africans to 1.3% in admixed Americans (**Figure 1d**).

Novel full- or partial-gene duplications were detected in 190 of 208 (91%) pharmacogenes studied. The most frequent were exonic duplications that affected *CBR3* and *CYP2B6* with 0.7% and 0.4% MAF in Finnish and African individuals, respectively (**Figure 1c**). Population-specific duplications included those in *SULT1A2* and *SLC13A1*, mainly present in South Asians and Africans (0.3% and 0.2% MAF, respectively), whereas duplications in *ABCC1* and *ABCC6* were detected in all populations with MAFs ranging between 0.1% and 0.2% (**Supplementary Table S1**). The aggregated frequency of the novel ADME gene duplications varied from 3.6% in South Asians to 2.1% in Finns (**Figure 1d**).

Table 1 Allele frequencies of previously established gene deletions in *GSTM1*, *GSTT1*, *UGT2B17*, *UGT2B28*, *CYP2A6*, *SULT1A1*, and *CYP2B6* genes

Population	Deletion MAF (%) ^a							
	<i>GSTM1</i> 1000 G [Reported]	<i>GSTT1</i> 1000 G [Reported]	<i>UGT2B17</i> 1000 G [Reported]	<i>UGT2B28</i> 1000 G [Reported]	<i>CYP2D6</i> 1000 G [Reported]	<i>CYP2A6</i> 1000 G [Reported]	<i>SULT1A1</i> 1000 G [Reported]	<i>CYP2B6</i> 1000 G [Reported]
Non-Finnish European	75 [77 (ref. 35)]	39 [39 (ref. 35)]	35 [30 (ref. 36)]	14 [13 (ref. 35)]	3 [3 (ref. 8)]	3 [2 (ref. 37)]	3	0
Finnish	75	38	42	23	2 [2 (ref. 38)]	2	1	0
East Asian	79 [70 (ref. 35)]	66 [65 (ref. 35)]	82 [82 (ref. 39)]	11 [15 (ref. 35)]	3 [7 (ref. 8)]	11 [8 (ref. 37)]	1	0
South Asian	57	45	61	26	3	5 [8 (ref. 37)]	2	0
African	49 [48 (ref. 35)]	55 [57 (ref. 35)]	24 [22 (ref. 35)]	34 [35 (ref. 35)]	6 [5 (ref. 8)]	3 [4 (ref. 37)]	1	0.5 [0.5 (ref. 37)]
Admixed American	69	49	35	22	3 [2 (ref. 40)]	4 [1 (ref. 37)]	1	0.1

MAF, minor allele frequency.

^aMAF derived from the 1000 Genomes Project phase 3 (1000 G) and, in brackets, previous reports in the literature. The deletions in the 1000 Genomes Project correspond to esv35587155, esv3647425, esv3600873, esv3600897, esv3647809, esv3644362, esv3638338, and esv3644370 for *GSTM1*, *GSTT1*, *UGT2B17*, *UGT2B28*, *CYP2D6*, *CYP2A6*, *SULT1A1*, and *CYP2B6*, respectively.

The newly described deletions encompassed genomic intervals with a median size of 24.7 kb, whereas duplications were slightly larger with a median size of 67.1 kb (**Figure 1e**). Overall, 87% of deletions and 93% of duplications exceeded sizes of 3 kb.

Gene deletions comprise a substantial fraction of ADME LOF alleles

Next, we analyzed the relative contributions of novel CNVs to the overall ADME LOF alleles. The contribution of novel deletions to the LOF alleles varied widely among the populations studied for many of the genes (**Figure 2**). While deletions of *CYP2C19* comprised 5.7% of all *CYP2C19* LOF alleles in Finns, the contribution was substantially lower (0.05–1%) in the other populations. Deletions of *CYP1A2* were identified exclusively in Africans with a low MAF (0.04%); however, they comprised around 57% of all *CYP1A2* LOF alleles in this population. In total, novel deletions accounted for > 5% of LOF alleles in a substantial number of genes (87, 25, 49, 48, 59 and 51 genes in non-Finnish Europeans, Finnish, East Asians, South Asians, Africans, and admixed Americans, respectively; **Figure 2**), emphasizing the overall importance of ADME CNVs.

In genes with previously described common CNVs, deletions constituted the majority of LOF alleles (**Supplementary Figure S2**). Yet, pronounced interethnic differences were evident for some of these genes, as illustrated by *CYP2D6*. The deletion allele *CYP2D6**5 accounted for 80% of all LOF alleles in East Asians, whereas it only accounted for 1–27% in the other populations studied.

Experimental validations of computational CNV predictions

To confirm the predictive quality of NGS-based CNV calls, we determined the breakpoints of three novel deletions overlapping *CYP2C19*, *CYP4F2*, and *SLCO1B3* (**Table 2**). The most frequent deletion in each of the genes was validated by PCR using primers predicted to amplify the deletion junctions. The precise breakpoints for the *CYP2C19* deletion, expanding from the promoter region to intron 5, *SLCO1B3* deletion from introns 8 to 13, and *CYP4F2* full-gene deletion affecting 13 exons, were located close to the computationally predicted deletion sites (**Figure 3**).

The *CYP2C19* exon 1–5 deletion was detected by genotyping with a MAF of 0.43% in a Finnish series of 465 subjects (0.8% MAF in this population according to ExAC; **Table 2**). The *SLCO1B3* exon 9–13 deletion was detected with 0.53% MAF in our 1,080 Spanish cohort (0.1% MAF in non-Finnish Europeans according to ExAC), and *CYP4F2* full-gene deletion was detected with a 1.61% MAF in 590 Japanese, including one subject homozygous for the gene deletion (0.4% MAF in East Asians according to ExAC).

DISCUSSION

The interindividual variability in drug pharmacokinetics and pharmacodynamics causes lack of drug efficacy and adverse drug reactions (ADRs) in a large fraction of patients.¹⁵ In

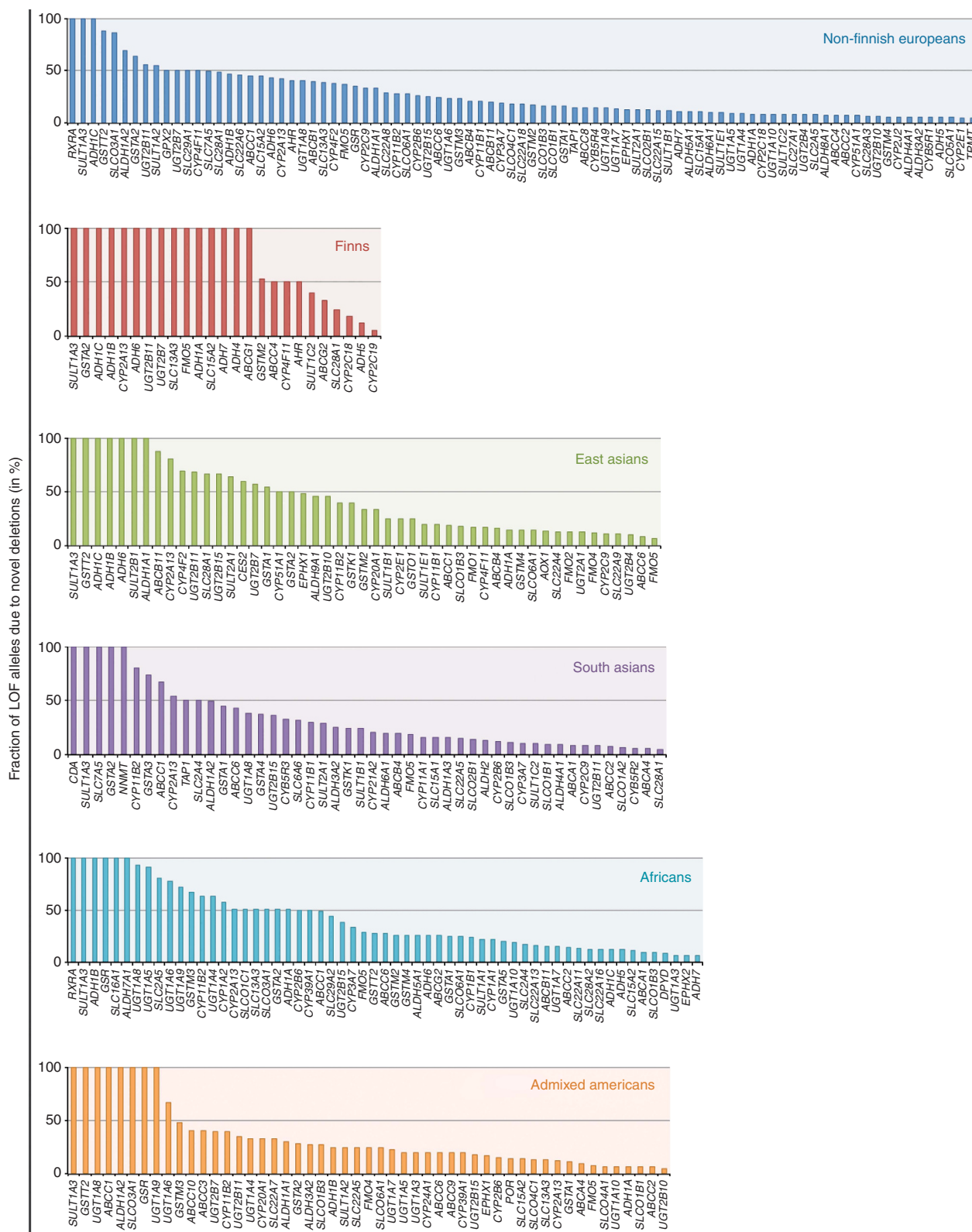


Figure 2 Fraction of loss-of-function (LOF) alleles attributable to novel deletions in six human populations. The relative contribution of novel deletions to the overall pool of LOF alleles is depicted for non-Finnish Europeans (blue), Finnish (red), East Asians (green), South Asians (purple), Africans (turquoise), and admixed Americans (orange). Only absorption, distribution, metabolism, and excretion (ADME) genes for which the contribution of novel deletions to LOF alleles exceeds 5% in the respective population are shown. Single-nucleotide variant (SNV) and indel LOF alleles were defined conservatively and include only those leading to frameshifts in the coding sequence, start-loss, stop-gain, or splice variants. Deletions of the *CYP2A7* pseudogene as well as previously described deletions are not included in this representation.

Table 2 Gene deletions in *CYP2C19*, *CYP4F2* and *SLCO1B3* described in ExAC (MAF > 0.05%)

Gene	Exons deleted (ExAC data)	Coordinates of deleted coding region ^a	MAF in ExAC		MAF by genotyping	
			(%)	Population ^b	(%)	Population (no. ^c)
<i>CYP2C19</i>	1–5	chr10:94,762,706–94,781,997	0.833	FIN	0.432 ^d	Finnish (463)
			0.091	NFE	0	Spanish (1,000)
			0	EAS	0	Japanese (590)
	6–7	chr10:94,820,496–94,843,024	0.163	EAS	Not tested	–
	1–6	chr10:94,762,706–94,820,637	0.091	FIN	Not tested	–
	2–5	chr10:94,775,058–94,781,997	0.076	FIN	Not tested	–
	1–5 ^e	chr10:94,735,263–94,781,997	0.061	FIN	Not tested	–
<i>SLCO1B3</i>	9–13	chr12:20,875,235–20,883,602	0.114	NFE	0.531 ^f	Spanish (1,035)
			0.023	EAS	0	Japanese (590)
			0.015	FIN	0	Finnish (462)
<i>CYP4F2</i>	2 ^g –13	chr19:15,897,611–15,878,771	0.363	EAS	1.610 ^h	Japanese (590)
			0.002	NFE	0	Spanish (1,065)
			0	FIN	0	Finnish (464)

ExAC, Exome Aggregation Consortium; MAF, minor allele frequency.

^aCoordinates refer to GRCh38 reference genome assembly. Transcripts correspond to ENST00000371321.7, ENST00000381545.7 and ENST00000221700.10. ^bFinnish (FIN), Non-Finnish Europeans (NFE) and East Asians (EAS). ^cNumber of successfully genotyped individuals. The total number of samples studied in each population are 465 Finnish, 1080 Spanish, 590 Japanese. ^dFour heterozygous subjects. ^eThis deletion ranges from *CYP2C18* exon 9 to *CYP2C19* exon 5. ^f11 heterozygous subjects. ^gExAC only provides data for coding regions. *CYP4F2* exon 1 is not translated and thus data for this region are not available in this database. ^h17 heterozygous and 1 homozygous deletion subjects.

total, ADRs cause around 6.5% of admissions to hospitals and can have severe or fatal outcomes especially in pediatric and geriatric patients.^{16,17} Furthermore, they account for 5–10% of annual hospital costs, posing an important economic burden on health-care services.¹⁸ Importantly, genetic factors are estimated to be responsible for 20–30% of observed ADRs and, thus, could be prevented by appropriated genetic tests.³ Accordingly, more than 100 drugs currently have pharmacogenomic labels to identify patients at risk for ADRs or lack of efficacy.^{19,20} However, these identified pharmacogenomic biomarkers only relate to frequent genetic variations, whereas recent large-scale sequencing projects have also revealed that rare variants are of relevance to drug pharmacokinetics.^{13,21,22} Besides being enriched in functional effects, overall rare alleles can be highly population-specific. One important example has been described for *CYP3A4*. While the LOF variant *CYP3A4*20* is globally rare, it is present at a high frequency in specific regions in Spain, where it significantly contributes to ADRs during paclitaxel therapy.^{23,24}

In recent years NGS techniques have burst into the field of genetics, providing novel and cost-effective tools to detect not only SNVs but also CNVs. NGS techniques suitable for the detection of CNVs have already been used in large research projects. Furthermore, the data presented here incentivizes the implementation of more comprehensive CNV testing also in the clinical setting once additional methodological requirements, particularly regarding analytic and clinical validity as well as the clinical utility, are met. Recently, CNV data from 59,898 human exomes was released, providing the largest and most powerful resource for the identification of this type of variation.¹¹ By leveraging this extensive NGS data set, we provide the first global panorama of the structural genetic diversity in pharmacogenes. Frequencies of most of the identified novel deletions varied

widely across populations and were estimated to substantially contribute to the total pool of LOF variants.

The gene with the highest number of novel deletions was *CYP2C19*, with a MAF in Finns of 1.1%. *CYP2C19* is involved in the metabolism of multitude of drugs and *CYP2C19* genotype-guided dosing recommendations are currently included in 21 Food and Drug Administration drug labels, including citalopram, clobazam, clopidogrel, escitalopram, and flibanserin (<https://www.fda.gov/>). We validated the most common deletion in *CYP2C19*, which we identified to affect exons 1–5, in a cohort of 465 Finns and found frequencies that were lower in the analyzed cohort than in computational predictions (predicted MAF = 0.8%, experimental MAF = 0.4%). This difference could be caused by the genotyping assay, which is specific for the breakpoints identified and would thus miss additional deletions bounded by alternative breakpoints but spanning the same exons. Regarding the contribution of *CYP2C19* deletions to the LOF phenotype, several frequent LOF single-nucleotide polymorphisms are known to affect this gene (e.g., *CYP2C19*2* and *CYP2C19*3* alleles); despite this, the deletion was estimated to represent 5.7% of all LOF alleles in the Finnish population.

In non-Finnish Europeans, deletions in *SLCO1B3* were the most common novel CNVs detected in the NGS data, accounting for 17% of all LOF alleles (Figure 2). The exon 9–13 deletion, predicted to have a 0.1% MAF in the general non-Finnish European population, occurred with a frequency of 0.5% in Spanish individuals. This 5-fold higher frequency in Spain might be caused by population-specific differences. OATP1B3, the transporter encoded by *SLCO1B3*, is important for the clearance of bilirubin²⁵ and has been implicated in the transport of multiple drugs, with substantially overlapping specificity with OATP1B1.²⁶ Importantly, however, some substrates including the angiotensin receptor blocker

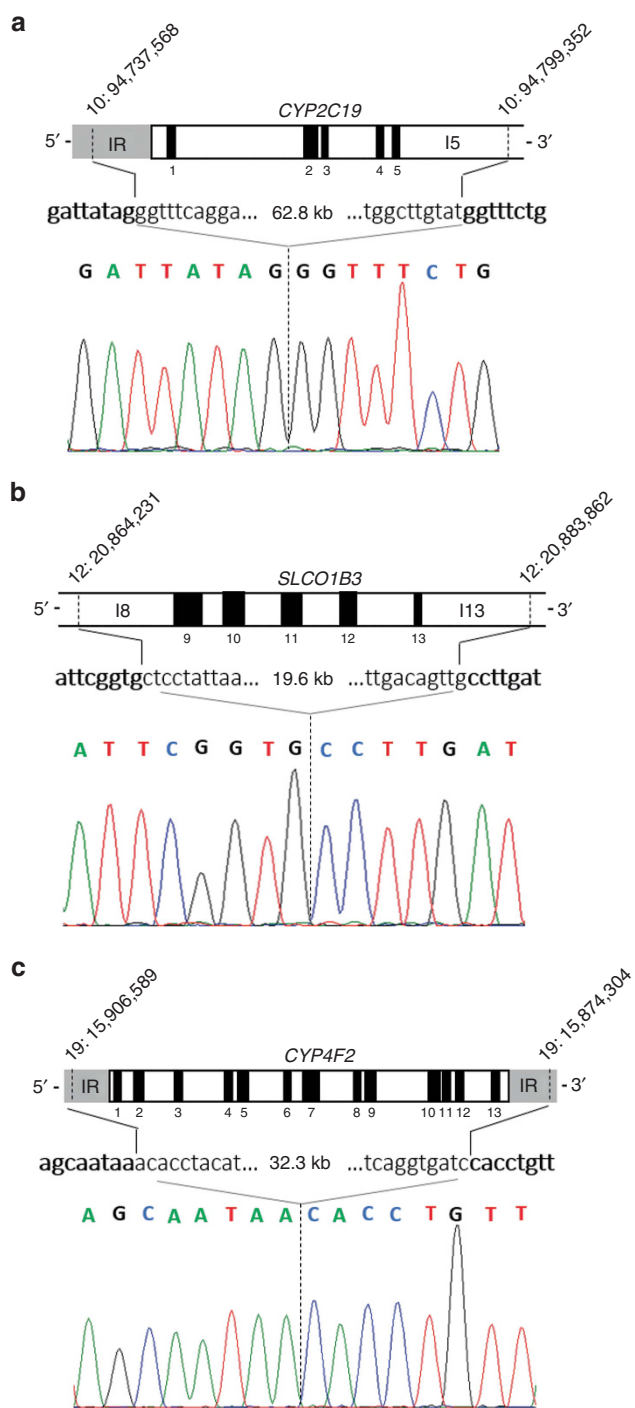


Figure 3 Experimental validation of newly identified *CYP2C19*, *SLCO1B3*, and *CYP4F2* deletions. Linear representation of the genomic reference sequence and the deletion breakpoints detected by Sanger sequencing in (a) *CYP2C19*, (b) *SLCO1B3*, and (c) *CYP4F2*. Sequences surrounding the deletion are shown in black and the deleted sequence is shown in gray. Coordinates refer to the GRCh38 reference genome assembly. IR, intergenic region.

telmisartan and the oncology drug docetaxel are transported primarily by OATP1B3.^{27–29}

Another relevant novel deletion spanned the *CYP4F2* gene, encoding a vitamin K oxidase that acts as an important

counterpart of VKORC1. Genetic polymorphisms in *VKORC1*, *CYP2C9*, and *CYP4F2* are clinically relevant for the dosing with anticoagulants, explaining up to 45% of warfarin dose requirements.^{30,31} The benefits of genotype-guided warfarin dosing on clinical outcomes have been studied in prospective clinical trials, albeit with sometimes conflicting results.^{32–34} *CYP4F2* deletions were found mainly in East Asians, and in this population they accounted for 69% of all LOF alleles. Genotyping for the novel *CYP4F2* full-gene deletion in Japanese revealed a subject homozygous for the deletion and a global MAF of 1.6%. The allele frequency found in this population was 4-fold higher than that predicted for East Asians (MAF = 0.36%) and might be explained by a population-specific distribution of the deletion. Thus, it can be suggested that genotyping for structural variants might further improve the selection of optimal anticoagulant starting doses and contribute to improved patient outcomes in East Asians.

In the era of precision medicine it is important to accurately characterize the genotypes of the specific patients subjected to pharmacotherapy. We here show that novel CNVs significantly contribute to the functionality of relevant pharmacogenes, adding an additional layer of pharmacogenetic complexity with important implications for the prediction of drug response and toxicity. We thus recommend the incorporation of CNV detection assays for relevant genes and populations. In combination with the important role of rare SNVs, our results suggest that the quality of preemptive pharmacogenetic advice, which is typically based on the interrogation of few candidate variants, can be improved by comprehensive NGS-based genotype identification of relevant pharmacogenes.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Economy and Competiveness (grant SAF2015-64850-R), by the European Union’s Horizon 2020 research and innovation program U-PGx under grant agreement 668353, and by the Swedish Research Council (grant agreements 2015-02760, 2016-01153, and 2016-01154).

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Lauschke VM, Ingelman-Sundberg M. The importance of patient-specific factors for hepatic drug response and toxicity. *Int J Mol Sci.* 2016;17: E1714.
2. Li J, Zhang L, Zhou H, Stoneking M, Tang K. Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum Mol Genet.* 2011;20:528–540.
3. Sim SC, Kacevska M, Ingelman-Sundberg M. Pharmacogenomics of drug-metabolizing enzymes: a recent update on clinical implications and endogenous effects. *Pharmacogenomics J.* 2013;13:1–11.

4. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337:64–69.
5. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291.
6. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444–454.
7. Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. *Annu Rev Genet*. 2011;45:203–226.
8. He Y, Hoskins JM, McLeod HL. Copy number variants in pharmacogenetic genes. *Trends Mol Med*. 2011;17:244–251.
9. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. *Genet Med*. 2017;19:69–76.
10. de Ligt J, Boone PM, Pfundt R, et al. Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat*. 2013;34:1439–1448.
11. Ruderfer DM, Hamamsy T, Lek M, et al. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat Genet*. 2016;48:1107–1111.
12. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
13. Kozyra M, Ingelman-Sundberg M, Lauschke VM. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet Med*. 2017;19:20–29.
14. Nakano M, Fukushima Y, Yokota S, et al. CYP2A7 pseudogene transcript affects CYP2A6 expression in human liver by acting as a decoy for miR-126. *Drug Metab Dispos*. 2015;43:703–712.
15. Spear BB, Heath-Chiozzi M, Huff J. Clinical application of pharmacogenetics. *Trends Mol Med*. 2001;7:201–204.
16. Impicciatore P, Choonara I, Clarkson A, Provasi D, Pandolfini C, Bonati M. Incidence of adverse drug reactions in paediatric in/out-patients: a systematic review and meta-analysis of prospective studies. *Br J Clin Pharmacol*. 2001;52:77–83.
17. Budnitz DS, Shehab N, Kegler SR, Richards CL. Medication use leading to emergency department visits for adverse drug events in older adults. *Ann Intern Med*. 2007;147:755–765.
18. Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient-episodes. *PLoS One*. 2009;4:e4439.
19. Frueh FW, Amur S, Mummaneni P, et al. Pharmacogenomic biomarker information in drug labels approved by the United States food and drug administration: prevalence of related drug use. *Pharmacotherapy*. 2008;28:992–998.
20. Ehmann F, Caneva L, Prasad K, et al. Pharmacogenomic information in drug labels: European Medicines Agency perspective. *Pharmacogenomics J*. 2015;15:201–210.
21. Lauschke VM, Ingelman-Sundberg M. Precision medicine and rare genetic variants. *Trends Pharmacol Sci*. 2016;37:85–86.
22. Lauschke VM, Ingelman-Sundberg M. Requirements for comprehensive pharmacogenetic genotyping platforms. *Pharmacogenomics*. 2016;17:917–924.
23. Apellaniz-Ruiz M, Lee MY, Sanchez-Barroso L, et al. Whole-exome sequencing reveals defective CYP3A4 variants predictive of paclitaxel dose-limiting neuropathy. *Clin Cancer Res*. 2015;21:322–328.
24. Apellaniz-Ruiz M, Inglada-Perez L, Naranjo ME, et al. High frequency and founder effect of the CYP3A4*20 loss-of-function allele in the Spanish population classifies CYP3A4 as a polymorphic enzyme. *Pharmacogenomics J*. 2015;15:288–292.
25. Sanna S, Busonero F, Maschio A, et al. Common variants in the SLCO1B3 locus are associated with bilirubin levels and unconjugated hyperbilirubinemia. *Hum Mol Genet*. 2009;18:2711–2718.
26. Kalliokoski A, Niemi M. Impact of OATP transporters on pharmacokinetics. *Br J Pharmacol*. 2009;158:693–705.
27. Ishiguro N, Maeda K, Kishimoto W, et al. Predominant contribution of OATP1B3 to the hepatic uptake of telmisartan, an angiotensin II receptor antagonist, in humans. *Drug Metab Dispos*. 2006;34:1109–1115.
28. Yamada A, Maeda K, Ishiguro N, et al. The impact of pharmacogenetics of metabolic enzymes and transporters on the pharmacokinetics of telmisartan in healthy volunteers. *Pharmacogenet Genomics*. 2011;21:523–530.
29. Yamada A, Maeda K, Kiyotani K, Mushihiro T, Nakamura Y, Sugiyama Y. Kinetic interpretation of the importance of OATP1B3 and MRP2 in docetaxel-induced hematopoietic toxicity. *CPT Pharmacometrics Syst Pharmacol*. 2014;3:e126.
30. Cooper GM, Johnson JA, Langae TY, et al. A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood*. 2008;112:1022–1027.
31. Johnson JA, Caudle KE, Gong L, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin Pharmacol Ther*. 2017;102:397–404.
32. Pirmohamed M, Burnside G, Eriksson N, et al. A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med*. 2013;369:2294–2303.
33. Kimmel SE, French B, Kasner SE, et al. A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med*. 2013;369:2283–2293.
34. Stergiopoulos K, Brown DL. Genotype-guided vs clinical dosing of warfarin and its analogues: meta-analysis of randomized clinical trials. *JAMA Intern Med*. 2014;174:1330–1338.
35. McCarroll SA, Hadnott TN, Perry GH, et al. Common deletion polymorphisms in the human genome. *Nat Genet*. 2006;38:86–92.
36. Gruber M, Le T, Filipits M, et al. UDP-glucuronosyltransferase 2B17 genotype and the risk of lung cancer among Austrian Caucasians. *Cancer Epidemiol*. 2013;37:625–628.
37. Martis S, Mei H, Vijzelaar R, Edelmann L, Desnick RJ, Scott SA. Multi-ethnic cytochrome-P450 copy number profiling: novel pharmacogenetic alleles and mechanism of copy number variation formation. *Pharmacogenomics J*. 2013;13:558–566.
38. Pietarinen P, Tornio A, Niemi M. High frequency of CYP2D6 ultrarapid metabolizer genotype in the Finnish population. *Basic Clin Pharmacol Toxicol*. 2016;119:291–296.
39. Okano M, Ueda T, Nishitani Y, Kano H, Ikekita A, Kageyama S. UDP-glucuronosyltransferase 2B17 genotyping in Japanese athletes and evaluation of the current sports drug testing for detecting testosterone misuse. *Drug Test Anal*. 2013;5:166–181.
40. Sosa-Macias M, Dorado P, Alanis-Banuelos RE, Llerena A, Lares-Asseff I. Influence of CYP2D6 deletion, multiplication, -1584C->G, 31G->A and 2988G->A gene polymorphisms on dextromethorphan metabolism among Mexican Tepehuanos and Mestizos. *Pharmacology*. 2010;86:30–36.