


## Article

# Combination of Genome-Wide Polymorphisms and Copy Number Variations of Pharmacogenes in Koreans

Nayoung Han, Jung Mi Oh and In-Wha Kim \* 

College of Pharmacy and Research Institute of Pharmaceutical Sciences,  
Seoul National University, Seoul 08826, Korea; hans1217@gmail.com (N.H.); jmoh@snu.ac.kr (J.M.O.)  
\* Correspondence: iwkim2@hanmail.net; Tel.: +82-2-880-7736

**Abstract:** For predicting phenotypes and executing precision medicine, combination analysis of single nucleotide variants (SNVs) genotyping with copy number variations (CNVs) is required. The aim of this study was to discover SNVs or common copy CNVs and examine the combined frequencies of SNVs and CNVs in pharmacogenes using the Korean genome and epidemiology study (KoGES), a consortium project. The genotypes ( $N = 72,299$ ) and CNV data ( $N = 1000$ ) were provided by the Korean National Institute of Health, Korea Centers for Disease Control and Prevention. The allele frequencies of SNVs, CNVs, and combined SNVs with CNVs were calculated and haplotype analysis was performed. *CYP2D6* rs1065852 (c.100C>T, p.P34S) was the most common variant allele (48.23%). A total of 8454 haplotype blocks in 18 pharmacogenes were estimated. *DMD* ranked the highest in frequency for gene gain (64.52%), while *TPMT* ranked the highest in frequency for gene loss (51.80%). Copy number gain of *CYP4F2* was observed in 22 subjects; 13 of those subjects were carriers with *CYP4F2*\*3 gain. In the case of *TPMT*, approximately one-half of the participants ( $N = 308$ ) had loss of the *TPMT*\*1\*1 diplotype. The frequencies of SNVs and CNVs in pharmacogenes were determined using the Korean cohort-based genome-wide association study.

**Keywords:** polymorphisms; pharmacogenes



**Citation:** Han, N.; Oh, J.M.; Kim, I.-W. Combination of Genome-Wide Polymorphisms and Copy Number Variations of Pharmacogenes in Koreans. *J. Pers. Med.* **2021**, *11*, 33. <https://doi.org/10.3390/jpm11010033>

Received: 13 November 2020  
Accepted: 4 January 2021  
Published: 7 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

It is well established that human genetic diversity is important for our understanding population histology [1], variability in disease susceptibility, and treatment response or adverse reactions to medications [2]. Single nucleotide variants (SNVs) are the most widely studied form of genetic variations and several SNVs have been linked to disease susceptibility and drug responses. Therefore, genome-wide association (GWA) studies have led to the identification of multiple genetic variants correlated with traits, such as body mass index, skin color [3], fat distribution [4], and glomerular filtration rate [5], and with diseases, such as autoimmune disease [6] and non-alcoholic fatty liver disease [7].

Additionally, these SNV markers from GWA studies can be used in pharmacogenomic research as a means of directly predicting interindividual responses to medicines [8]. Research has identified successfully the loci of genetic variants associated with responses to tumor necrosis factor inhibitors [9], antidepressants [10], and antipsychotics [11], and with adverse reactions induced by medicines, such as thiopurine-induced myelosuppression [12], statin-induced myopathy [13], and carbamazepine-induced hypersensitivity [14]. These genetic variations alter the structure and function of proteins such as drug-metabolizing enzymes, drug transporters, receptors, and response targets, collectively referred to as pharmacogenes [15].

Common copy number variations (CNVs) were estimated to occur in approximately 9.5% of the human reference genome and have non-random distribution [16]. CNVs account for at least five times more variable base pairs compared to that of SNVs when two human genomes are compared to each other [17,18]. As with SNVs, CNVs were found to

influence susceptibility to cancer [19] as well as neurodegenerative disease [20] and psychiatric disease [21]. Despite their clinical significance, CNVs remain understudied compared to SNVs. The reasons may be that the detection of CNVs is more difficult and CNVs only occur with low-to-intermediate frequency [22]. However, for predicting phenotypes and executing precision medicine, combination analysis of SNVs genotyping with CNVs is required. There have been several studies to detect both CNVs and SNVs in *CYP2D6* [23,24]. However, CNV information integrated with polymorphisms on pharmacogenes is still not fully characterized [25]. Traditional methods are time-consuming and labor-intensive and a large number of participants are required.

The Korean genome and epidemiology study (KoGES) is a consortium project that was established as a genome epidemiological study for the research community with a health database and biobank to help investigate Korean population-based and gene–environment model studies [26–28]. Because this dataset contains a significant collection of SNVs and CNVs data from normal tissue and blood samples, KoGES is appropriate for combined pharmacogenomic studies. Thus, this study aimed to discover SNVs and CNVs and to examine the combined frequencies of SNVs and CNVs in pharmacogenes in the Korean population using this large public dataset.

## 2. Materials and Methods

### 2.1. Study Subjects

The study subjects were selected from the Ansan and Ansong study ( $N = 5836$ ), the Health Examinee cohort (HEXA,  $N = 58,701$ ), and the cardiovascular disease association study (CAVAS,  $N = 8105$ ) that constitute the KoGES [29]. Epidemiologic data were provided by the Korean National Institute of Health, Korea Centers for Disease Control and Prevention (KCDC). Socio-demographic, medical history, health conditions, and family history of disease information were collected by trained interviewers using structured questionnaires. All physical examinations were administered by health professionals trained to follow standardized protocols. The participants who had cancer were excluded from the analysis. All subjects were middle-aged adults between 40 and 69 years of age. All study participants provided written informed consent.

### 2.2. Pharmacogenes

The pharmacogenomics-related genes were selected by the Very Important Pharmacogene summaries in the Pharmacogenomics Knowledge Base (as of March 2020) [30] and the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline (as of March 2020) [31]. The genes from the U.S. Food and Drug Administration (FDA) Table of Pharmacogenomic Biomarkers in Drug Labels (as of March 2020) were included [32]. A total of 191 genes were analyzed and are listed in Supplementary Table S1.

### 2.3. Data Collection and Preprocessing

The genotypes ( $N = 72,299$ ) and CNV data ( $N = 1000$ ) were provided by the KCDC. These imputed genotypes were produced by the Korea BioBank Array (referred to as KoreanChip, KCHIP, Seoul, The Republic of Korea) project, optimized for the Korean population [33]. A KCHIP array includes a total of 833,535 SNVs for autosomal chromosomes. Quality-controlled data were used for imputation analysis with 1000 Genomes Phase 3 data as a reference panel using ShapeIT v2 [34] and IMPUTE v2 [35]. An SNV missing rate greater than 0.05, SNVs with a minor allele frequency less than 0.01, or a Hardy–Weinberg equilibrium (HWE) of  $P$  less than  $10^{-6}$  were excluded according to standard quality control procedures. The SNV position aligned to human reference genomes hg19 using the Bioconductor BiomaRt R package [36]. For each gene, 10 kb bases of region were added both upstream and downstream of the defined gene location. The CNV data of 1000 subjects were produced from the Ansan and Ansong study [37]. The CNV data were genotyped with the NimbleGen HD2  $3 \times 720$  K comparative genomic hybridization array (aCGH) (Roche NimbleGen, Madison, WI, USA) [37]. For the combination analysis of genotypes

and CNVs, the variants from gene–drug pairs from CPIC were searched for their clinical effects. The functional effects of variants were predicted by SIFT (Sorting Intolerance From Tolerant) [38] and POLYPHEN-2 (Polymorphism Phenotyping v2) [39].

#### 2.4. CNV Calling

R package that implements the Genome Alteration Detection Analysis algorithm (GADA) was used for CNV discovery [40]. To overcome the limitation of single algorithm detection, we tested different thresholds,  $T$ , from 3 to 8. CNV discovery with several parameters was tested to find the best parameters using known CNV regions [41]. Consequently, we selected the best parameter with high concordance with known CNV regions with  $T = 4.5$ ,  $\alpha = 0.2$ , and  $\text{MinSegLen} = 6$ . CNV regions longer than 50 bp in length were included for further analysis. A log 2 ratio cut-off of  $\pm 0.25$  was used to define copy number gain and loss and cut-offs of  $\pm 0.8$  were used to define amplification and deletion, respectively [42,43].

#### 2.5. Data Analysis

Categorical variables such as gender and variant occurrences are presented in percentages and frequencies. Continuous variables such as age are presented with average and standard variations. The chi-squared test with one degree of freedom was used to test the departure from HWE for each variant. Data were analyzed with PLINK 1.9 or 2.0 [44] and R (version 3.6.3). Linkage disequilibrium analysis among pairs of SNVs was performed to identify the haplotype. Estimation of haplotype blocks and their frequencies were performed with PLINK and Haploview [45].

### 3. Results

#### 3.1. Characteristics of the Study Population

For the KCHIP study, among the Ansan and Ansung study ( $N = 5493$ ), HEXA ( $N = 58,701$ ), and CAVAS ( $N = 8105$ ), after excluding patients with cancer, 5182 of the Ansan and Ansung study subjects, 55,955 of HEXA, and 7890 of CAVAS remained. For the CNV data, 945 subjects remained after excluding patients with cancer. Among them, 614 subjects had SNV and CNV data. The characteristics of the subjects from the SNV and CNV data are presented in Table 1. The average ages of the subjects with SNV and CNV data were 54.08 and 54.05 years, respectively. The frequencies of female subjects (63.78%) was higher than that of male subjects (36.22%) in the SNV data, while that of female subjects (49.95%) was similar to that of male subjects (50.05%) in the CNV data.

**Table 1.** Demographic characteristics of study subjects.

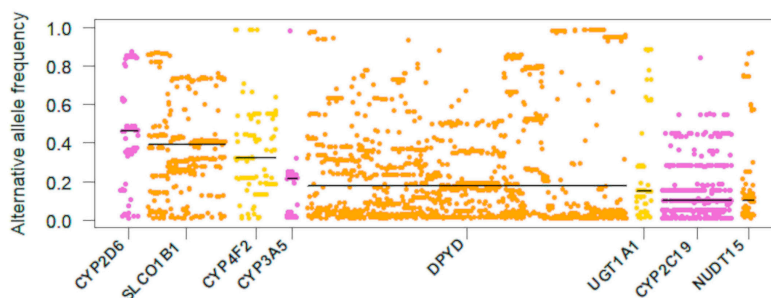
Characteristics	SNV	CNV	Combination of SNV with CNV
Number of patients, $n$	69,027	947	614
Age, years	54.08 $\pm$ 8.31	54.05 $\pm$ 9.08	52.82 $\pm$ 8.80
Gender			
male	25,004 (36.22)	474 (50.05)	311 (50.65)
female	44,023 (63.78)	473 (49.95)	303 (49.35)

Values are reported as  $n$  (%) or mean  $\pm$  standard deviation; SNV, single nucleotide variation; CNV, copy number variation.

#### 3.2. Genotype Variants

A total of 36,853 SNVs in pharmacogenes were included for the further analysis. The allele frequencies of SNVs of more than 10% are listed in Supplementary Table S2. *VKORC1* rs9923231 (−1639G>A or G3673A) was found to be the most common alternative allele (92.42%). *CYP2D6* rs1065852 (c.100C>T, p.P34S) was the next common allele (48.23%). The allelic frequencies of *CYP2C19\*2* (rs4244285, c.681G>A, p.P227P) and *CYP2C19\*3* (rs4986893, c.636G>A, p.W212X) were 28.29% and 10.04%, respectively. The allelic frequency of *CYP3A5\*3* (rs776746, c.6986A>G) was 23.47%. *CYP4F2\*3* (rs2108622,

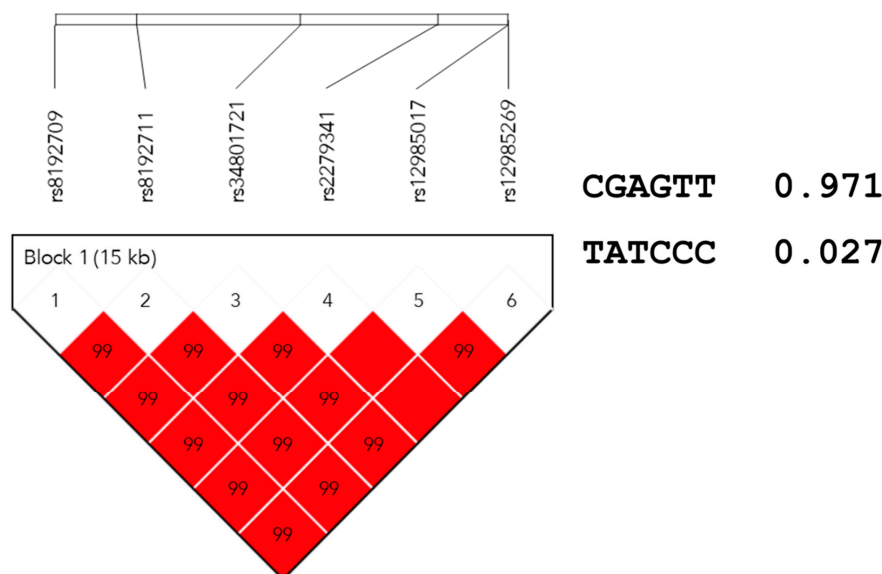
c.1297C>T, p.V433M) and *CYP4F2*\*2 rs3093105 (c.34T>G, p.W12G) were 32.41% and 13.40%, respectively. Among SNVs in pharmacogenes, those that were assigned as having level A evidence of gene–drug pairs by CPIC are shown in Figure 1. The median alternative allele frequency of *CYP2D6* variants was ranked the highest (46.17%, ranged from 1.02% to 87.34%), followed by *SLCO1B1* variants (39.32%, ranged from 1.07% to 86.62%). SNVs with frequencies less than 10% that were also assigned as having level A evidence of gene–drug pairs by CPIC or predicted to be deleterious by SIFT and POLYPHEN-2 are listed in Supplementary Table S3. *CACNA1S* rs3850625 (c.4615G>A (p.R1539C), *CFTR* rs121909046 (c.650A>G, p.E217G) and rs113857788 (c.4056G>C p.Q1352H), and *CYP2B6* rs8192709 (c.64C>T, p.R22C) were predicted to be deleterious by SIFT.



**Figure 1.** Single-nucleotide variants of pharmacogenes with alternative allele frequencies of more than 10% in a Korean population. Horizontal lines indicate median values.

### 3.3. Haplotype Analysis

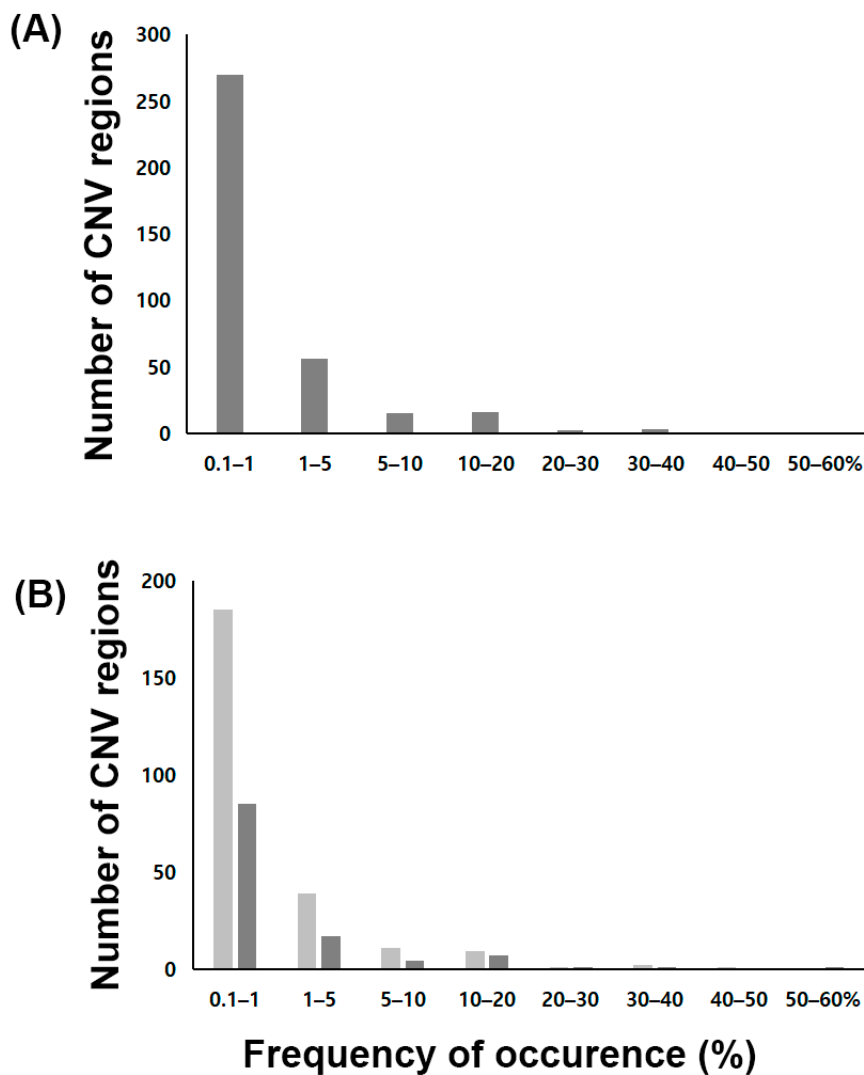
The frequency distributions of the variants or haplotypes were found to be significantly different among ethnic populations. Therefore, haplotype analysis was performed on about 18 pharmacogenomic genes from 73 gene–drug pairs with level A evidence by CPIC. A total of 8454 haplotype blocks in 18 genes were estimated, and the number varied from 2 to 3924 blocks per each gene, with an average of 4378. *CYP2B6* rs8192709 (c.64C>T, p.R22C) constructed a haplotype block with rs8192711 (G>A), rs34801721 (A>T), rs2279341 (G>C), rs12985017 (T>C), and rs12985269 (T>C) (Figure 2). The haplotype block of *CYP2B6* in Caucasians was constructed with rs2279341, rs12985017, and rs12985269. Carriers with the alternative haplotype T-A-T-C-C were found in 3.98% of this study population.



**Figure 2.** Haplotype block map constructed by candidate single-nucleotide variations in *CYP2B6*. Notes: Block 1 includes rs8192709, rs8192711, rs34801721, rs2279341, rs12985017, and rs12985269; the linkage disequilibrium between two SNPs is indicated by the standardized  $r^2$  (red boxes).

### 3.4. Copy Number Variation Profiling

In the 947 subjects, segments with more than 1 CNV were determined in 937 subjects using GADA. In total, 448 segments were detected in 937 individuals with an average of 22.58 copy number segments in each individual. CNV regions of more than 50 bp were included for the further analysis. The mean and median lengths of these CNV regions were 4.29 and 2.21 kb, respectively. Figure 3 shows the distribution of the 333 CNV regions by frequency rate. Of the 333 CNV regions, 92 had frequency rates of >1%. The frequencies of CNVs were calculated and genes with a frequency of more than 1% are summarized in Table 2. *DMD* ranked the highest in frequency for gene gain (64.52%), while *TPMT* ranked the highest in frequency for gene loss (51.80%), and the frequency of *TPMT* deletion was 3.58%. There were gene gains in *G6PD* (17.21%), *KIT* (21.12%), and *OTC* (57.76%), while there was gene loss in *ABCB1* (15.31%), *BCR* (19.01%), *DMD* (20.27%), *EGFR* (41.39%), *HLA-B* (36.54%), *HLA-DRB1* (40.65%), *PDGFRA* (21.44%), and *SULT1A1* (19.75%) with a frequency of more than 10%. The genes with a CNV frequency of less than 1% are listed in Supplementary Table S4. Gene losses of *ABCG2* and *CYP2E1* were found in 0.63% of subjects, while the gene gain of *CYP2B6* was found in 0.21% of subjects.



**Figure 3.** Distribution of copy number variation frequencies for the copy number variation regions in a Korean population. (A) Frequencies of copy number variation regions. (B) Copy number variation frequencies of the detected copy number variation regions, divided into gains and losses.

**Table 2.** Copy number variations for pharmacogenes with a frequency of more than 1% in Koreans.

Gene	Position	Gain Frequency (%)	Loss Frequency (%)
<i>ABCB1</i>	7: 87133179–87342639	0.11	15.31
<i>ALK</i>	2: 29415640–30144477	6.12	1.06
<i>ALOX5</i>	10: 45869624–45941567	6.65	1.58
<i>BCR</i>	11: 23522552–23660224	0.11	19.01
<i>BRCA</i>	17: 41196312–41277500	2.22	2.64
<i>COMT</i>	19: 19929263–19957498	7.07	0.32
<i>CYP2A6</i>	19: 41349443–41356352	1.27	1.48
<i>CYP4F2</i>	19: 15988834–16008884	3.80	0.42
<i>DMD</i>	X: 31137345–33229673	64.52	20.27
<i>EGFR</i>	7: 55086725–55275031	2.32	41.39
<i>ESR1</i>	6: 152128814–152424408	0	1.48
<i>G6PD</i>	X: 153759606–153775233	17.21	0.42
<i>HLA-B</i>	6: 31237743–31324989	0.42	36.54
<i>HLA-DRB1</i>	6: 32489683–32557613	0.32	40.65
<i>KIT</i>	4: 55524095–55606881	21.12	2.22
<i>OTC</i>	X: 38211736–38280703	57.76	0.42
<i>PDGFRA</i>	4: 55095264–55164412	0.11	21.44
<i>RYR1</i>	19: 38924340–39078204	5.07	2.11
<i>SMN2</i>	5: 70220768–70248842	2.11	0.11
<i>SULT1A1</i>	16: 28616908–28620649	7.71	19.75
<i>TPMT</i>	6: 18128545–18155374	0	51.80

### 3.5. Combination of Genotype Variants and CNVs

A total of 22 pharmacogenomic genes from 73 gene–drug pairs with level A evidence by CPIC were selected for the combination analysis of SNVs and CNVs in 614 subjects. *CYP4F2*\*1\*3 (24.43%) was most common *CYP4F2* diplotype followed by *CYP4F2*\*2\*3 (18.57%) (Table 3). Among the *CYP4F2* gains observed in 22 subjects, 13 subjects were carriers with a *CYP4F2*\*3 gain. The frequency of *CYP4F2* loss was 0.49%. In the *TPMT* case, approximately half of the participants ( $N = 308$ ) showed a loss of the *TPMT*\*1\*1 diplotype.

**Table 3.** Copy number variation combined with single nucleotide variations in Koreans ( $N = 614$ ).

Gene Allele	Subjects (N)	Frequency (%)
<i>CYP4F2</i> *1*1	258	42.02
<i>CYP4F2</i> *1*2	1	0.16
<i>CYP4F2</i> *1*3	150	24.43
<i>CYP4F2</i> *3*3	22	3.58
<i>CYP4F2</i> *2*3	114	18.57
<i>CYP4F2</i> *1*2-*3*3	31	5.05
<i>CYP4F2</i> *2*2-*3*3	13	2.12
<i>CYP4F2</i> *1*1 gain	9	1.47
<i>CYP4F2</i> *1*3 gain	6	0.98
<i>CYP4F2</i> *2*3 gain	6	0.98
<i>CYP4F2</i> *3*3 gain	1	0.16
<i>CYP4F2</i> *1*1 loss	2	0.33
<i>CYP4F2</i> *2*3 loss	1	0.16
<i>TPMT</i> *1*1	287	46.74
<i>TPMT</i> *1*3C	8	1.30
<i>TPMT</i> *1*1 loss	308	50.16
<i>TPMT</i> *1*3C loss	11	1.79

## 4. Discussion

Pharmacogenomic studies represent a critical component of precision medicine. Compared to SNVs, CNVs or the combined study of SNVs and CNVs all both relatively less studied. With regard to SNV or CNV data from genome epidemiological research,

KoGES in Korea can be used for pharmacogenomic studies. The purpose of this study was to discover SNVs and CNVs and to examine the combined frequencies of SNVs and CNVs in pharmacogenes in Korea using KoGES.

For 191 pharmacogenes, a total of 36,853 SNVs from 69,027 subjects, 333 CNVs from 947 subjects, and combined data of SNVs and CNVs from 614 subjects were available in this study. The SNV rs9923231 (−1639G>A or G3673A) is known to alter a transcription factor binding site in the *VKORC1* promoter region, and this allele frequency in Asians was found to be approximately 0.92 [46], similar with our result. This variant was associated with decreased gene expression, resulting in decreased warfarin dose requirements. *CYP2D6* rs1065852 (c.100C>T, p.P34S) was the next most common allele (48.23%), and it appeared in \*4, \*10, \*14A, and \*36 alleles, with lower enzyme activities compared to the wild type [47]. This enzyme is involved in the metabolism of approximately 25% of commonly prescribed drugs, including antidepressants, antipsychotics, antiarrhythmics,  $\beta$ -blockers, and opioids [24]. The allelic frequencies of *CYP2C19*\*2 and *CYP2C19*\*3 were 28.29% and 10.04%, respectively, in our study, similar to earlier findings [48], and indicate that genomic data from the KoGES study are appropriate for pharmacogenomics studies in Korea. These losses of functional alleles of *CYP2C19* can increase the risks for serious cardiovascular events among patients treated with clopidogrel [49].

In our study, *CACNA1S* rs3850625 and *CFTR* rs121909046 and rs113857788 were predicted to be deleterious by SIFT. *CACNA1S* rs3850625 was associated with malignant hyperthermia accelerated by inhalational anesthetics and muscle relaxants [50]. Those two variants in the *CFTR* gene were found to have the strongest association with bronchiectasis and chronic pancreatitis in the Korean population [51].

According to a haplotype analysis, the haplotype block *CYP2B6*\*2 (rs8192709) was constructed and the corresponding frequency was found to be 3.98 in this study. Approximately 3.4% of *CYP2B6*\*2 variants were found in Han and Uygur Chinese [52]. Although the level of evidence for clinical annotations of *CYP2B6*\*2 was lower than that for the *CYP2B6*\*6 allele according to CPIC, this minor allele is known to decrease the clearance of methadone [53] or efavirenz [54].

The activities of several important drug-metabolizing genes, such as *CYP2B6*, *CYP2E1*, *CYP2D6*, *GSTM1*, and *SULT1A1*, are known to be related to variable copy numbers. In our study, CNVs of *CYP2B6*, *CYP2E1*, and *SULT1A1* were detected, whereas CNV data from KoGES did not cover *CYP2D6* and *GSTM1* genes. Accordingly, alternative methods during a CNV analysis are needed to detect those genes.

The *DMD* gene found to be the most frequent in terms of the copy number gain in our study is the largest gene in the human genome, encompassing 2.2 Mb and encoding for a muscular protein, dystrophin, which is related to the X-linked recessive disorders Duchenne muscular dystrophy and Becker muscular dystrophy [55]. Deletions or complex rearrangements usually occur between exons 43 and 55 or exons 2 and 23 [55]. Most carriers with mutations or deletions of the *DMD* gene are asymptomatic [56]. One hundred and seventeen different deletions and 48 duplications in the *DMD* gene were found in 507 Korean patients with Duchenne muscular dystrophy or Becker muscular dystrophy [57].

*TPMT* ranked highest here in terms of the frequency of gene loss at 3.58% in our results. This is most likely due to a variable number of tandem repeats (VNTR) within a G/C-rich region in the promoter of *TPMT* [58]. The frequency of the VNTR allele, consisting of two repeat sequence motifs A, one motif B, and one motif C, was reported to be 48.2% in an Asian British cohort [59]. The patterns and total number of VNTR alleles were associated with the level of TPMT activity [59]. The *TPMT* gene encoding thiopurine S-methyltransferase is a crucial enzyme during the metabolism of thiopurine drugs such as azathioprine and 6-mercaptopurine [60].

In the next step, the CNV data were combined with SNVs for pharmacogenes. The loss frequency of *TPMT*\*3C (rs1142345, c.A719G, p.Y240C) was 1.79%. The *TPMT*\*3C variant, with moderate activity, was found to be the most frequent alternative allele in Koreans, and TPMT deficiency can increase certain fatal adverse reaction risks, such as bone marrow

toxicity and myelosuppression induced by 6-mercaptopurine [61]. Thiopurine-associated leukopenia (more than 30%) was found to be considerably higher than expected according to the frequency of the *TPMT* variant (~1%) in Koreans with Crohn's disease [62]. This result may be related to the copy number variation in the *TPMT* gene. Despite the fact that less than 5% of the samples showed gene gains or losses in these genes, the corresponding clinical impacts should be considered when medicines associated with these genes are administered.

A limitation of this study was that the CNV frequencies of some genes differed from those in previous studies [63]. This difference may have been caused by the different assay methods. There are many different methods for determining the CNVs of genes, and each method has advantages and pitfalls. The array CGH methods and SNP arrays and CNV arrays are excellent for initial scans along the lines of the SNP GWA study, and other PCR-based methods such as multiplex ligation-dependent probe amplification (MLPA) are used for conformation to genotype copy numbers [64]. The KCHIP array did not contain SNVs for sex chromosomes, meaning that pharmacogenes such as *DMD* and *G6PD* could not be included in the analysis of the combinations of genotype variants and CNVs. Another limitation in our study was that hybrid pseudogene, conversion, or tandem alleles cannot be determined due to the assay method used in this study. Additionally, as subjects with common complex diseases such as diabetic mellitus, hypertension, and cardiovascular disease were not excluded, this could affect the results of this study. Further studies with regard to functional variation evaluations and associated determinations are needed to manage patients more efficiently.

The 1000 Genomes Project and the Encyclopedia of DNA Elements Project produced comprehensive maps outlining the regions of the human genome containing SNVs, multi-nucleotide variants, and CNVs [65]. However, combination analyses of SNVs with CNVs in pharmacogenetic studies are limited. Here, we conducted a combined analysis of SNVs with CNVs in pharmacogenes in Koreans.

In conclusion, the frequencies of SNVs and CNVs in pharmacogenes were determined by means of a Korean cohort-based GWA study. Though further assessments of correlations with phenotype changes are necessary, the results here may be useful for the identification of genetic causes of cases involving severe drug-induced toxicity or reduced therapeutic benefits from a drug.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2075-4426/11/1/33/s1>, Table S1: List of 129 pharmacogenes in this study. Table S2: Single nucleotide variants of pharmacogenes with an allele frequency of more than 10% in Koreans. Table S3: Single nucleotide variants of pharmacogenes with an allele frequency of more than 10% in Koreans. Table S4. Copy number variations for pharmacogenes with a frequency of less than 1% in Koreans.

**Author Contributions:** Conceptualization, I.-W.K. and J.M.O.; methodology, N.H.; validation, I.-W.K.; formal analysis, N.H.; investigation, N.H.; data curation, N.H.; writing—original draft preparation, N.H.; writing—review and editing, I.-W.K.; visualization, N.H.; supervision, I.-W.K. and J.M.O.; funding acquisition, I.-W.K. and J.M.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the National Research Foundation of Korea grant funded by the Korean government (MSIT) (No. 2018R1A2B6001859 and 2019R1A2C1005211).

**Institutional Review Board Statement:** This study was exempt from the institutional review board of Seoul National University (IRB No. E1910/001-002).

**Informed Consent Statement:** The requirement for written informed consent from participants was waived because all participants were anonymized by National Biobank of Korea, the Centers for Disease Control and Prevention, Republic of Korea.

**Data Availability Statement:** The datasets generated during the current study are available from the corresponding authors on reasonable request.



**Acknowledgments:** This study was conducted with bioresources from the National Biobank of Korea, the Centers for Disease Control and Prevention, Republic of Korea (KBN-2019-065).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**, *536*, 285–291. [[CrossRef](#)] [[PubMed](#)]
2. Zhang, H.; Liu, R.; Yan, C.; Liu, L.; Tong, Z.; Jiang, W.; Yao, M.; Fang, W.; Chen, Z. Advantage of next-generation sequencing in dynamic monitoring of circulating tumor DNA over droplet digital PCR in cetuximab treated colorectal cancer patients. *Transl. Oncol.* **2019**, *12*, 426–431. [[CrossRef](#)] [[PubMed](#)]
3. Galvan-Femenia, I.; Obon-Santacana, M.; Pineyro, D.; Guindo-Martinez, M.; Duran, X.; Carreras, A.; Pluvinet, R.; Velasco, J.; Ramos, L.; Ausso, S.; et al. Multitrait genome association analysis identifies new susceptibility genes for human anthropometric variation in the GCAT cohort. *J. Med. Genet.* **2018**, *55*, 765–778. [[CrossRef](#)] [[PubMed](#)]
4. Tachmazidou, I.; Suveges, D.; Min, J.L.; Ritchie, G.R.S.; Steinberg, J.; Walter, K.; Iotchkova, V.; Schwartzentruber, J.; Huang, J.; Memari, Y.; et al. Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. *Am. J. Hum. Genet.* **2017**, *100*, 865–884. [[CrossRef](#)] [[PubMed](#)]
5. Hellwege, J.N.; Velez Edwards, D.R.; Giri, A.; Qiu, C.; Park, J.; Torstenson, E.S.; Keaton, J.M.; Wilson, O.D.; Robinson-Cohen, C.; Chung, C.P.; et al. Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. *Nat. Commun.* **2019**, *10*, 3842. [[CrossRef](#)]
6. Din, L.; Sheikh, M.; Kosaraju, N.; Smedby, K.E.; Bernatsky, S.; Berndt, S.I.; Skibola, C.F.; Nieters, A.; Wang, S.; McKay, J.D.; et al. Genetic overlap between autoimmune diseases and non-Hodgkin lymphoma subtypes. *Genet. Epidemiol.* **2019**, *43*, 844–863. [[CrossRef](#)]
7. Namjou, B.; Lingren, T.; Huang, Y.; Parameswaran, S.; Cobb, B.L.; Stanaway, I.B.; Connolly, J.J.; Mentch, F.D.; Benoit, B.; Niu, X.; et al. GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE Network. *BMC Med.* **2019**, *17*, 135. [[CrossRef](#)] [[PubMed](#)]
8. Popejoy, A.B. Diversity in precision medicine and pharmacogenetics: Methodological and conceptual considerations for broadening participation. *Pharmgenom. Pers. Med.* **2019**, *12*, 257–271. [[CrossRef](#)]
9. Massey, J.; Plant, D.; Hyrich, K.; Morgan, A.W.; Wilson, A.G.; Spiliopoulou, A.; Colombo, M.; McKeigue, P.; Isaacs, J.; Cordell, H.; et al. Genome-wide association study of response to tumour necrosis factor inhibitor therapy in rheumatoid arthritis. *Pharmacogenom. J.* **2018**, *18*, 657–664. [[CrossRef](#)] [[PubMed](#)]
10. Fabbri, C.; Tansey, K.E.; Perlis, R.H.; Hauser, J.; Henigsberg, N.; Maier, W.; Mors, O.; Placentino, A.; Rietschel, M.; Souery, D.; et al. New insights into the pharmacogenomics of antidepressant response from the GENDEP and STAR\*D studies: Rare variant analysis and high-density imputation. *Pharmacogenom. J.* **2018**, *18*, 413–421. [[CrossRef](#)]
11. Yu, H.; Yan, H.; Wang, L.; Li, J.; Tan, L.; Deng, W.; Chen, Q.; Yang, G.; Zhang, F.; Lu, T.; et al. Five novel loci associated with antipsychotic treatment response in patients with schizophrenia: A genome-wide association study. *Lancet Psychiatry* **2018**, *5*, 327–338. [[CrossRef](#)]
12. Walker, G.J.; Harrison, J.W.; Heap, G.A.; Voskuil, M.D.; Andersen, V.; Anderson, C.A.; Ananthakrishnan, A.N.; Barrett, J.C.; Beaugerie, L.; Bewshea, C.M.; et al. Association of genetic variants in NUDT15 with thiopurine-induced myelosuppression in patients with inflammatory bowel disease. *JAMA* **2019**, *321*, 773–785. [[CrossRef](#)]
13. Carr, D.F.; Francis, B.; Jorgensen, A.L.; Zhang, E.; Chinoy, H.; Heckbert, S.R.; Bis, J.C.; Brody, J.A.; Floyd, J.S.; Psaty, B.M.; et al. Genomewide association study of statin-induced myopathy in patients recruited using the UK clinical practice research datalink. *Clin. Pharmacol. Ther.* **2019**, *106*, 1353–1361. [[CrossRef](#)] [[PubMed](#)]
14. Nicoletti, P.; Barrett, S.; McEvoy, L.; Daly, A.K.; Aithal, G.; Lucena, M.I.; Andrade, R.J.; Wadelius, M.; Hallberg, P.; Stephens, C.; et al. Shared genetic risk factors across carbamazepine-induced hypersensitivity reactions. *Clin. Pharmacol. Ther.* **2019**, *106*, 1028–1036. [[CrossRef](#)]
15. Ahmed, S.; Zhou, Z.; Zhou, J.; Chen, S.Q. Pharmacogenomics of drug metabolizing enzymes and transporters: Relevance to precision medicine. *Genom. Proteom. Bioinform.* **2016**, *14*, 298–313. [[CrossRef](#)]
16. Zarrei, M.; MacDonald, J.R.; Merico, D.; Scherer, S.W. A copy number variation map of the human genome. *Nat. Rev. Genet.* **2015**, *16*, 172–183. [[CrossRef](#)]
17. Pang, A.W.; MacDonald, J.R.; Pinto, D.; Wei, J.; Rafiq, M.A.; Conrad, D.F.; Park, H.; Hurles, M.E.; Lee, C.; Venter, J.C.; et al. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol.* **2010**, *11*, R52. [[CrossRef](#)] [[PubMed](#)]
18. Sudmant, P.H.; Mallick, S.; Nelson, B.J.; Hormozdiari, F.; Krumm, N.; Huddleston, J.; Coe, B.P.; Baker, C.; Nordenfelt, S.; Bamshad, M.; et al. Global diversity, population stratification, and selection of human copy-number variation. *Science* **2015**, *349*, aab3761. [[CrossRef](#)] [[PubMed](#)]
19. Arcella, A.; Limanaqi, F.; Ferese, R.; Biagioni, F.; Oliva, M.A.; Storto, M.; Fanelli, M.; Gambardella, S.; Fornai, F. Dissecting molecular features of gliomas: Genetic loci and validated biomarkers. *Int. J. Mol. Sci.* **2020**, *21*, 685. [[CrossRef](#)] [[PubMed](#)]
20. Gentile, G.; La Cognata, V.; Cavallaro, S. The contribution of CNVs to the most common aging-related neurodegenerative diseases. *Aging Clin. Exp. Res.* **2020**. [[CrossRef](#)] [[PubMed](#)]

21. Sullivan, P.F.; Owen, M.J. Increasing the clinical psychiatric knowledge base about pathogenic copy number variation. *Am. J. Psychiatry* **2020**, *177*, 204–209. [[CrossRef](#)] [[PubMed](#)]
22. Lauer, S.; Gresham, D. An evolving view of copy number variants. *Curr. Genet.* **2019**, *65*, 1287–1295. [[CrossRef](#)]
23. Kim, J.; Lee, S.Y.; Lee, K.A. Copy number variation and gene rearrangements in CYP2D6 genotyping using multiplex ligation-dependent probe amplification in Koreans. *Pharmacogenomics* **2012**, *13*, 963–973. [[CrossRef](#)]
24. Qiao, W.; Martis, S.; Mendiratta, G.; Shi, L.; Botton, M.R.; Yang, Y.; Gaedigk, A.; Vijzelaar, R.; Edelman, L.; Kornreich, R.; et al. Integrated CYP2D6 interrogation for multiethnic copy number and tandem allele detection. *Pharmacogenomics* **2019**, *20*, 9–20. [[CrossRef](#)] [[PubMed](#)]
25. Santos, M.; Niemi, M.; Hiratsuka, M.; Kumondai, M.; Ingelman-Sundberg, M.; Lauschke, V.M.; Rodriguez-Antona, C. Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genet. Med.* **2018**, *20*, 622–629. [[CrossRef](#)] [[PubMed](#)]
26. Han, J.; Shon, J.; Hwang, J.Y.; Park, Y.J. Effects of Coffee Intake on Dyslipidemia Risk According to Genetic Variants in the ADORA Gene Family among Korean Adults. *Nutrients* **2020**, *12*, 493. [[CrossRef](#)] [[PubMed](#)]
27. Kwon, Y.J.; Hong, K.W.; Park, B.J.; Jung, D.H. Serotonin receptor 3B polymorphisms are associated with type 2 diabetes: The Korean Genome and Epidemiology Study. *Diabetes Res. Clin. Pract.* **2019**, *153*, 76–85. [[CrossRef](#)]
28. Yang, Y.J.; Kim, J.; Kwock, C.K. Association of Genetic Variation in the Epithelial Sodium Channel Gene with Urinary Sodium Excretion and Blood Pressure. *Nutrients* **2018**, *10*, 612. [[CrossRef](#)]
29. Kim, Y.; Han, B.G. Cohort Profile: The Korean Genome and Epidemiology Study (KoGES) Consortium. *Int. J. Epidemiol.* **2017**, *46*, 1350. [[CrossRef](#)]
30. Thorn, C.F.; Klein, T.E.; Altman, R.B. PharmGKB: The Pharmacogenomics Knowledge Base. *Methods Mol. Biol.* **2013**, *1015*, 311–320.
31. Relling, M.V.; Klein, T.E. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin. Pharmacol. Ther.* **2011**, *89*, 464–467. [[CrossRef](#)] [[PubMed](#)]
32. Tutton, R. Pharmacogenomic biomarkers in drug labels: What do they tell us? *Pharmacogenomics* **2014**, *15*, 297–304. [[CrossRef](#)] [[PubMed](#)]
33. Moon, S.; Kim, Y.J.; Han, S.; Hwang, M.Y.; Shin, D.M.; Park, M.Y.; Lu, Y.; Yoon, K.; Jang, H.M.; Kim, Y.K.; et al. The Korea Biobank array: Design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* **2019**, *9*, 1382. [[CrossRef](#)] [[PubMed](#)]
34. Delaneau, O.; Marchini, J.; Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **2011**, *9*, 179–181. [[CrossRef](#)]
35. Marchini, J.; Howie, B.; Myers, S.; McVean, G.; Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **2007**, *39*, 906–913. [[CrossRef](#)]
36. Durinck, S.; Moreau, Y.; Kasprzyk, A.; Davis, S.; De Moor, B.; Brazma, A.; Huber, W. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **2005**, *21*, 3439–3440. [[CrossRef](#)]
37. Moon, S.; Kim, Y.J.; Hong, C.B.; Kim, D.J.; Lee, J.Y.; Kim, B.J. Data-driven approach to detect common copy-number variations and frequency profiles in a population-based Korean cohort. *Eur. J. Hum. Genet.* **2011**, *19*, 1167–1172. [[CrossRef](#)]
38. Sim, N.L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P.C. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic. Acids. Res.* **2012**, *40*, W452–W457. [[CrossRef](#)]
39. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249. [[CrossRef](#)]
40. Pique-Regi, R.; Caceres, A.; Gonzalez, J.R. R-Gada: A fast and flexible pipeline for copy number analysis in association studies. *BMC Bioinform.* **2010**, *11*, 380. [[CrossRef](#)]
41. Moon, S.; Kim, Y.J.; Kim, Y.K.; Kim, D.J.; Lee, J.Y.; Go, M.J.; Shin, Y.A.; Hong, C.B.; Kim, B.J. Genome-wide survey of copy number variants associated with blood pressure and body mass index in a Korean population. *Genom. Inform.* **2011**, *9*, 152–160. [[CrossRef](#)]
42. Bambury, R.M.; Bhatt, A.S.; Riester, M.; Peadamallu, C.S.; Duke, F.; Bellmunt, J.; Stack, E.C.; Werner, L.; Park, R.; Iyer, G.; et al. DNA copy number analysis of metastatic urothelial carcinoma with comparison to primary tumors. *BMC Cancer* **2015**, *15*, 242. [[CrossRef](#)] [[PubMed](#)]
43. Lindgren, D.; Sjodahl, G.; Lauss, M.; Staaf, J.; Chebil, G.; Lovgren, K.; Gudjonsson, S.; Liedberg, F.; Patschan, O.; Mansson, W.; et al. Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. *PLoS ONE* **2012**, *7*, e38863. [[CrossRef](#)]
44. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **2015**, *4*, 7. [[CrossRef](#)]
45. Barrett, J.C.; Fry, B.; Maller, J.; Daly, M.J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**, *21*, 263–265. [[CrossRef](#)]
46. Owen, R.P.; Gong, L.; Sagreiya, H.; Klein, T.E.; Altman, R.B. VKORC1 pharmacogenomics summary. *Pharm. Genom.* **2010**, *20*, 642–644. [[CrossRef](#)]
47. Sakuyama, K.; Sasaki, T.; Ujiie, S.; Obata, K.; Mizugaki, M.; Ishikawa, M.; Hiratsuka, M. Functional characterization of 17 CYP2D6 allelic variants (CYP2D6.2, 10, 14A-B, 18, 27, 36, 39, 47-51, 53-55, and 57). *Drug Metab. Dispos.* **2008**, *36*, 2460–2467. [[CrossRef](#)]

48. Kim, K.A.; Song, W.K.; Kim, K.R.; Park, J.Y. Assessment of CYP2C19 genetic polymorphisms in a Korean population using a simultaneous multiplex pyrosequencing method to simultaneously detect the CYP2C19\*2, CYP2C19\*3, and CYP2C19\*17 alleles. *J. Clin. Pharm. Ther.* **2010**, *35*, 697–703. [[CrossRef](#)]
49. Scott, S.A.; Sangkuhl, K.; Stein, C.M.; Hulot, J.S.; Mega, J.L.; Roden, D.M.; Klein, T.E.; Sabatine, M.S.; Johnson, J.A.; Shuldiner, A.R.; et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin. Pharmacol. Ther.* **2013**, *94*, 317–323. [[CrossRef](#)]
50. Carpenter, D.; Ringrose, C.; Leo, V.; Morris, A.; Robinson, R.L.; Halsall, P.J.; Hopkins, P.M.; Shaw, M.A. The role of CACNA1S in predisposition to malignant hyperthermia. *BMC Med. Genet.* **2009**, *10*, 104. [[CrossRef](#)]
51. Lee, J.H.; Choi, J.H.; Namkung, W.; Hanrahan, J.W.; Chang, J.; Song, S.Y.; Park, S.W.; Kim, D.S.; Yoon, J.H.; Suh, Y.; et al. A haplotype-based molecular analysis of CFTR mutations associated with respiratory and pancreatic diseases. *Hum. Mol. Genet.* **2003**, *12*, 2321–2332. [[CrossRef](#)] [[PubMed](#)]
52. Guan, S.; Huang, M.; Li, X.; Chen, X.; Chan, E.; Zhou, S.F. Intra- and inter-ethnic differences in the allele frequencies of cytochrome P450 2B6 gene in Chinese. *Pharm. Res.* **2006**, *23*, 1983–1990. [[CrossRef](#)]
53. Gadel, S.; Crafford, A.; Regina, K.; Kharasch, E.D. Methadone N-demethylation by the common CYP2B6 allelic variant CYP2B6.6. *Drug Metab. Dispos.* **2013**, *41*, 709–713. [[CrossRef](#)]
54. Paganotti, G.M.; Russo, G.; Sobze, M.S.; Mayaka, G.B.; Muthoga, C.W.; Tawe, L.; Martinelli, A.; Romano, R.; Vullo, V. CYP2B6 poor metaboliser alleles involved in efavirenz and nevirapine metabolism: CYP2B6\*9 and CYP2B6\*18 distribution in HIV-exposed subjects from Dschang, Western Cameroon. *Infect. Genet. Evol.* **2015**, *35*, 122–126. [[CrossRef](#)]
55. Ishmukhametova, A.; Chen, J.M.; Bernard, R.; de Massy, B.; Baudat, F.; Boyer, A.; Mechin, D.; Thorel, D.; Chabrol, B.; Vincent, M.C.; et al. Dissecting the structure and mechanism of a complex duplication-triplication rearrangement in the DMD gene. *Hum. Mutat.* **2013**, *34*, 1080–1084. [[CrossRef](#)]
56. Iskandar, K.; Dwianingsih, E.K.; Pratiwi, L.; Kalim, A.S.; Mardhiah, H.; Putranti, A.H.; Nurputra, D.K.; Triono, A.; Herini, E.S.; Malueka, R.G.; et al. The analysis of DMD gene deletions by multiplex PCR in Indonesian DMD/BMD patients: The era of personalized medicine. *BMC Res. Notes* **2019**, *12*, 704. [[CrossRef](#)]
57. Cho, A.; Seong, M.W.; Lim, B.C.; Lee, H.J.; Byeon, J.H.; Kim, S.S.; Kim, S.Y.; Choi, S.A.; Wong, A.L.; Lee, J.; et al. Consecutive Analysis of Mutation Spectrum in the Dystrophin Gene of 507 Korean Boys with Duchenne/Becker Muscular Dystrophy in a Single Center. *Muscle Nerve* **2017**, *55*, 727–734. [[CrossRef](#)]
58. Spire-Vayron de la Moureyre, C.; Debuysere, H.; Fazio, F.; Sergent, E.; Bernard, C.; Sabbagh, N.; Marez, D.; Lo Guidice, J.M.; D'Halluin, J.C. Characterization of a variable number tandem repeat region in the thiopurine S-methyltransferase gene promoter. *Pharmacogenetics* **1999**, *9*, 189–198.
59. Urbancic, D.; Smid, A.; Stocco, G.; Decorti, G.; Mlinaric-Rascan, I.; Karas Kuzelicki, N. Novel motif of variable number of tandem repeats in TPMT promoter region and evolutionary association of variable number of tandem repeats with TPMT\*3 alleles. *Pharmacogenomics* **2018**, *19*, 1311–1322. [[CrossRef](#)]
60. Green, D.J.; Duong, S.Q.; Burckart, G.J.; Sissung, T.; Price, D.K.; Figg, W.D., Jr.; Brooks, M.M.; Chinnock, R.; Canter, C.; Addonizio, L.; et al. Association Between Thiopurine S-Methyltransferase (TPMT) Genetic Variants and Infection in Pediatric Heart Transplant Recipients Treated With Azathioprine: A Multi-Institutional Analysis. *J. Pediatr. Pharmacol. Ther.* **2018**, *23*, 106–110. [[CrossRef](#)]
61. Lee, S.S.; Kim, W.Y.; Jang, Y.J.; Shin, J.G. Duplex pyrosequencing of the TPMT\*3C and TPMT\*6 alleles in Korean and Vietnamese populations. *Clin. Chim. Acta* **2008**, *398*, 82–85. [[CrossRef](#)] [[PubMed](#)]
62. Yang, S.K.; Hong, M.; Baek, J.; Choi, H.; Zhao, W.; Jung, Y.; Haritunians, T.; Ye, B.D.; Kim, K.J.; Park, S.H.; et al. A common missense variant in NUDT15 confers susceptibility to thiopurine-induced leukopenia. *Nat. Genet.* **2014**, *46*, 1017–1020. [[CrossRef](#)] [[PubMed](#)]
63. Vijzelaar, R.; Botton, M.R.; Stolk, L.; Martis, S.; Desnick, R.J.; Scott, S.A. Multi-ethnic SULT1A1 copy number profiling with multiplex ligation-dependent probe amplification. *Pharmacogenomics* **2018**, *19*, 761–770. [[CrossRef](#)] [[PubMed](#)]
64. Guthrie, P.A.; Gaunt, T.R.; Abdollahi, M.R.; Rodriguez, S.; Lawlor, D.A.; Smith, G.D.; Day, I.N. Amplification ratio control system for copy number variation genotyping. *Nucleic. Acids. Res.* **2011**, *39*, e54. [[CrossRef](#)] [[PubMed](#)]
65. Haraksingh, R.R.; Snyder, M.P. Impacts of variation in the human genome on gene regulation. *J. Mol. Biol.* **2013**, *425*, 3970–3977. [[CrossRef](#)]