PLOS ONE

RESEARCH ARTICLE

Analysis of Pharmacogenomic Variants Associated with Population Differentiation

Bora Yeon[1☯], Eunyong Ahn[1☯], Kyung-Im Kim[2], In-Wha Kim[2], Jung Mi Oh[2], Taesung Park[1,3*]

1 Interdisciplinary Program in Bioinformatics, Seoul National University, Gwanak-ro, Gwanak-gu, Seoul, Korea, **2** College of Pharmacy, Seoul National University, Gwanak-ro, Gwanak-gu, Seoul, Korea, **3** Department of Statistics, Seoul National University, Seoul, Korea

☯ These authors contributed equally to this work.
* tspark@stats.snu.ac.kr

Citation: Yeon B, Ahn E, Kim KI, Kim IW, Oh JM, Park T (2015) Analysis of Pharmacogenomic Variants Associated with Population Differentiation. PLoS ONE 10(3): e0119994. doi:10.1371/journal.pone.0119994

Academic Editor: Nicholas John Timpson, University of Bristol, UNITED KINGDOM

Received: July 2, 2014

Accepted: February 3, 2015

Published: March 25, 2015

Data Availability Statement: All relevant data are within the manuscript, from the HapMap Phase III database (http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/hapmap3_r3/), and from the 'genes.zip' download link on the Pharmacogenomics Knowledge Base database download site (https://www.pharmgkb.org/downloads/).

Funding: This study was supported by the National Research Foundation of Korea (NRF) grants funded by the Republic of Korea (MSIP) (2012R1A3A2026438, 2008–0062618, 2013M3A9C4078158). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

In the present study, we systematically investigated population differentiation of drug-related (DR) genes in order to identify common genetic features underlying population-specific responses to drugs. To do so, we used the International HapMap project release 27 Data and Pharmacogenomics Knowledge Base (PharmGKB) database. First, we compared four measures for assessing population differentiation: the chi-square test, the analysis of variance (ANOVA) F-test, F_{st}, and Nearest Shrunken Centroid Method (NSCM). F_{st} showed high sensitivity with stable specificity among varying sample sizes; thus, we selected F_{st} for determining population differentiation. Second, we divided DR genes from PharmGKB into two groups based on the degree of population differentiation as assessed by F_{st}: genes with a high level of differentiation (HD gene group) and genes with a low level of differentiation (LD gene group). Last, we conducted a gene ontology (GO) analysis and pathway analysis. Using all genes in the human genome as the background, the GO analysis and pathway analysis of the HD genes identified terms related to cell communication. "Cell communication" and "cell-cell signaling" had the lowest Benjamini-Hochberg's q-values (0.0002 and 0.0006, respectively), and "drug binding" was highly enriched (16.51) despite its relatively high q-value (0.0142). Among the 17 genes related to cell communication identified in the HD gene group, five genes (*STX4*, *PPARD*, *DCK*, *GRIK4*, and *DRD3*) contained single nucleotide polymorphisms with F_{st} values greater than 0.5. Specifically, the F_{st} values for rs10871454, rs6922548, rs3775289, rs1954787, and rs167771 were 0.682, 0.620, 0.573, 0.531, and 0.510, respectively. In the analysis using DR genes as the background, the HD gene group contained six significant terms. Five were related to reproduction, and one was "Wnt signaling pathway," which has been implicated in cancer. Our analysis suggests that the HD gene group from PharmGKB is associated with cell communication and drug binding.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

The DNA sequence of the 3-billion-nucleotide-long human genome varies by approximately 0.1% between individuals. Surprisingly, this small difference in the DNA sequence accounts for individual differences in appearance, behavior, and even disease status. Furthermore, this difference in DNA sequence can have an even larger effect among ethnic populations. Genetic divergence between ethnic groups is called population differentiation (PD). PD results from genetic factors that enforce natural selection, genetic drift, or gene flow. Moreover, genes related to Mendelian diseases have a significant excess of single-nucleotide polymorphisms (SNPs) with high levels of PD, and the incidence of and susceptibility to these diseases differ among populations [1].

Several recent studies on PD have focused on genetic variations. Myles *et al.* attempted to identify SNPs accounting for disease-associated PD [2]. However, they found no disease-associated SNPs that were more significantly differentiated than randomly selected SNPs in the genome among populations. Nevertheless, the frequencies of risk alleles for disease-associated SNPs showed substantial variation across human populations. Barreiro *et al.* analyzed the degree of PD with 2.8 million SNPs and discovered the role of natural selection in shaping PD [3]. Wu and Zhang also performed a genome-wide study of PD and found that many groups of genes had higher degrees of PD [1]. Specifically, PD existed on some loci associated with phenotypes (e.g., hair growth and pigmentation) that are well known to vary across populations.

PD has also been investigated in pharmacogenomic studies [4]. For example, the response to warfarin, one of the most widely studied drugs, depends not only on genetic variants [5] but also on population [6]. As a result, some authors have suggested that warfarin be dosed according to the patient's race. In fact, Pavani *et al.* suggested a linear model for optimizing population-specific warfarin dose [7]. Huang *et al.* identified SNPs contributing to etoposide-induced cytotoxicity in a genome-wide association study (GWAS) using International HapMap cell lines, and they demonstrated different genotypes associated with cytotoxicity between two populations [8]. In order to investigate PD of DR genes, we analyzed data from two databases: International HapMap release 27 (phase II + III) [9] and Pharmacogenomics Knowledge Base (PharmGKB) [10,11], the most widely used DR database. Originally, HapMap release 27 contained 11 subpopulations. However, the allele frequencies of populations in the same ethnic groups are highly correlated [12], and there is lack of genotypic information in some populations. Therefore, we used the following subpopulations: European, African, and Asian from Japan and China.

There are several measures for determining the distance among populations. Among them, F_{st} is the most widely used measure to determine PD. Akey *et al.* [13] and Barreiro *et al.* [3] used Weir and Cockerham's estimate, an unbiased estimate of F_{st} [14,15]. Casto *et al.* used four measures: (i) δ, the difference in allele frequency between two groups; (ii) integrated haplotype score (iHS), which characterizes the lengths of the haplotypes surrounding each allele of a SNP [16]; (iii) latitude/longitude correlation (LLC), which describes how closely changes in a SNP's allele frequency follow geographical coordinates; and (iv) F_{st}, which shows variation in allele frequency among populations [17]. Park *et al.* used the Nearest Shrunken Centroid Method (NSCM) [18,19], which was originally designed for clustering of microarray data. NSCM has been proposed for solving the classification problem with a large number of features and it was also applied to the analysis of population differentiation in SNPs via Hapmap data [18]. Han *et al.* modified F_{st} for use with allele frequency data with unbalanced sample sizes [20].

In order to investigate PD of DR genes, we first compared four measures for assessing population differentiation: the chi-square test, the ANOVA F-test, F_{st}, and NSCM. F_{st} showed high sensitivity with stable specificity among varying sample sizes; thus, we selected F_{st} for determining population differentiation. We then divided DR genes from PharmGKB into two

groups based on the degree of population differentiation as assessed by F_{st}: genes with high a level of differentiation (HD gene group) and genes with a low level of differentiation (LD gene group). Finally, we conducted a gene ontology (GO) analysis and pathway analysis.

Several studies have investigated PD associated with individual drugs [4]. In the present study, we systematically studied PD of drug-related (DR) genes by simultaneously considering all reported DR genes. This integrative approach may help clarify the inconsistent genetic features of drug response associated with PD. Furthermore, our findings will improve the study and prediction of drug responses that differ among populations due to genetic stratification.

Methods

1. Measures of PD

Since the measures of PD are not always consistent, it is difficult to choose an appropriate measure for PD. Thus, we first performed a comparison analysis in order to identify the highest performing measures in our study. We compared the following four measures: Weir and Cockerham's F_{st} [15], the sum of square of d_i from NSCM, the chi-square test, and analysis of variance (ANOVA) F-test. Other measures were excluded for the following reasons. δ is used for comparisons between two populations; however, we compared PD among three populations. In our research, we tried to evaluate which SNPs are highly differentiated but iHS shows whether SNPs are differently selected. Therefore, the results via iHS are not concordant with the results from other measures. Moreover, Ferrer-Admetlla et al. suggest that iHS seems to be affected by the recombination rate [21]. Thus, we would like to exclude iHS from our sensitivity and specificity analysis. LLC was excluded, because latitude and longitude information for each individual was needed to determine PD.

We compared the specificity and sensitivity of these measures using simulation studies. Our comparison study focused on consistency and reliability with respect to the populations' sample sizes and imbalance in sample sizes among populations. Our comparison revealed that F_{st} had the most stable specificity regardless of the variability in sample size and the highest sensitivity as compared to other measures. Thus, we concluded that F_{st} was the most appropriate measure of PD for our integrative analysis of International HapMap release 27 and PharmGKB.

The chi-square test is a widely used statistical method for testing the homogeneity of group proportions. In this study, we used it to test whether allele frequencies of the *J* subgroups were equal; the null hypothesis was:

$$p_1 = \cdots = p_J \tag{1}$$

where p_i denotes the allele frequency of the i_{th} population. In the chi-square test, 0.05 or the q-value from Benjamini and Hochberg's method [22] is usually used as the significance level for testing the null hypothesis. Thus, the significance level varies according to *N*.

The model for the ANOVA F-test was:

$$a_{ij} = \mu + \tau_i + \grave{o}_{ij}, \sum \tau_i = 0 \tag{2}$$

where a_{ij} is the number of the allele (value of 0, 1, or 2) for the j_{th} individual in the i_{th} population. μ and $\mu+\tau_i$ are the overall mean genotype frequencies within individuals and mean of allele frequencies in the i_{th} population, respectively. ϵ_{ij} is the error term. Thus, by testing H_0: $\tau_i = 0$, for $^{\forall}i$, we could test whether the allele frequencies of subgroups were equal to one another assuming a Gaussian distribution.

We used Weir's F_{st} estimate $\hat{\theta}$ [14,15], an unbiased estimator of F_{st}. n_i denotes the sample size of the i_{th} subpopulation (i = 1,. . .,*s*). n = Σn_i denotes the total sample size. $\hat{p}_i$denotes the

observed allele frequency of the i_{th} subpopulation, and $\bar{p} = \sum n_i \hat{p}_i / n$ denotes the weighted average of allele frequency.

$$\hat{\theta} = \frac{\text{MSP} - \text{MSG}}{\text{MSP} + (\text{n}_\text{c} - 1)\text{MSG}}, \tag{3}$$

where

$$\text{MSP} = \frac{1}{s-1} \sum_{i=1}^{s} n_i (\hat{p}_i - \bar{p})^2, \tag{4}$$

$$\text{MSG} = \frac{1}{\sum_{i=1}^{s} (n_i - 1)} \sum_{i=1}^{s} n_i \hat{p}_i (1 - \hat{p}_i), \tag{5}$$

$$n_c = \frac{1}{s-1} \left(\sum_{i=1}^{s} n_i - \frac{\sum_{i=1}^{s} n_i^2}{\sum_{i=1}^{s} n_i} \right) \tag{6}$$

MSP and MSG represent the observed mean square error of a locus between populations and the observed mean square error of a locus within populations, respectively. n_c is the average sample size across n samples, correcting for variation in sample size among subpopulations.

We also defined the sum of square of standardized distance to measure PD via NSCM as follows;

$$\text{SS}_\text{d} = \sum_i d_{ik}^2 \tag{7}$$

It is a representative value for the k_{th} SNP in population i, where

$$d_{ik} = \frac{a_{ik} - a_k}{m_i (s_0 + s_k)} \tag{8}$$

Here, a_{ik} denotes the mean of allele frequencies in population i; a_k denotes the overall mean of allele frequency of SNP k, and $m_i = \sqrt{\frac{1}{n_k} + \frac{1}{n}}$, which makes $m_i \cdot s_k$ equal to the estimate of standard error for the numerator of $d_{ik} \cdot s_0$ was set equal to the median of s_k over the set of SNPs to prevent inflation of d_{ik}.

2. GO analysis and pathway analysis of the HD and LD gene groups

We used a gene ontology (GO) analysis to identify biological characteristics of the HD and LD gene groups. We compared each gene group to other functionally annotated genes in HapMap Data [23] and to DR genes in the PharmGKB Database.

Wright proposed the following F_{st} categories: (i) $F_{st} < 0.05$, low divergence; (ii) $0.05 < F_{st} < 0.15$, moderate divergence; (iii) $0.15 < F_{st} < 0.25$, high divergence; and (iv) $F_{st} > 0.25$, very high divergence [24]. Using Wright's F_{st} criteria, genes containing at least one SNP with an F_{st} value greater than 0.25 were considered to have a high level of differentiation (HD gene group) [25,26], while those containing SNPs with F_{st} values less than 0.05 were considered to have a low level of differentiation (LD gene group). Additionally, we identified the SNPs with a high level of differentiation from GO analysis results if F_{st} greater than 0.5, because this criterion was used for previous studies [1,27].

For the GO analysis, SNPs associated with drugs from PharmGKB were annotated into genes. These DR genes were divided into two groups using the F_{st} criteria proposed by Wright [24]. From 654 DR SNPs in the HapMap Database, we obtained 160 SNPs with HD and 173 SNPs with LD (Table 1). From these SNPs, 68 genes with HD and 114 genes with LD were derived.

To investigate the biological differences between the HD and LD gene groups, we performed a GO analysis and a pathway analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [28] v6.7 functional annotation tool. Annotated genes from each group were used as the input, while a list of whole genes in DAVID with at least one annotation in the analyzing categories was used as the background. For the GO analysis, the following three categories were selected: biological process (BP), molecular function (MF), and cellular component (CC) [29]. For the pathway analysis, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway was used [30].

Additional GO and pathway analyses were performed in a similar manner in order to compare genes in the HD gene group to those in the DR gene group. In this case, the DR HD gene group was used as the input for analysis, and the DR gene group was used as the background.

To correct for multiple tests, we used the hypergeometric test from Benjamini-Hochberg's method [22]. Fold enrichments, defined as the ratios of proportions between the input and background, were calculated for each term. Terms with Benjamini-Hochberg's q-values of 0.05 or lower were considered significant.

Results

1. Data collection

We analyzed SNP data from International HapMap Phase II + III, release 27 (http://www.hapmap.org) [9,31]. According to Hapmap consortium, there are distinct three clusters, European, African and Asian from the principal component plot of 11 populations in hapmap3 [31]. Therefore, we used three groups based on these three regions. We included 120 Yoruba from Ibadan, Nigeria (YRI), 181 Asians of which are 91 Japanese from Tokyo, Japan (JPT) and 90 Han Chinese from Beijing, China (CHB), and 120 Utah residents with ancestry from northern and western Europe (CEU). We used only founders of CEU and YRI to exclude the related samples. Because International HapMap release 27 consists of a mixture of two phases, each SNP had a different sample size. Fig. 1 shows the sample-size distributions of subpopulations from International HapMap Data. The SNPs from Phase II had smaller sample sizes, while those from Phase III had larger sample sizes (Fig. 1B). Some SNPs are only genotyped in phase II and others are only genotyped in phase III. In this reason, we only included four populations, which are both in phase II, and III simultaneously to avoid the potential biases due to different settings of each phases.

In addition, Kim et al. (2012) reported that JPT and CHB in Hapmap possess the common genetic information through MDS plot and Fst [32]. Therefore, we merged these two data into one East Asian data. We used allele frequency data of three populations as following: Yoruba in Ibadan, Nigeria (YRI), Utah residents with ancestry from Northern and Western Europe

Table 1. Summary of each SNP group.

	SNPs with high differentiation	SNPs with low differentiation	Total
Count	160	173	654
Mean (Fst)	0.364	0.020	0.157
Median (Fst)	0.344	0.019	0.111

doi:10.1371/journal.pone.0119994.t001

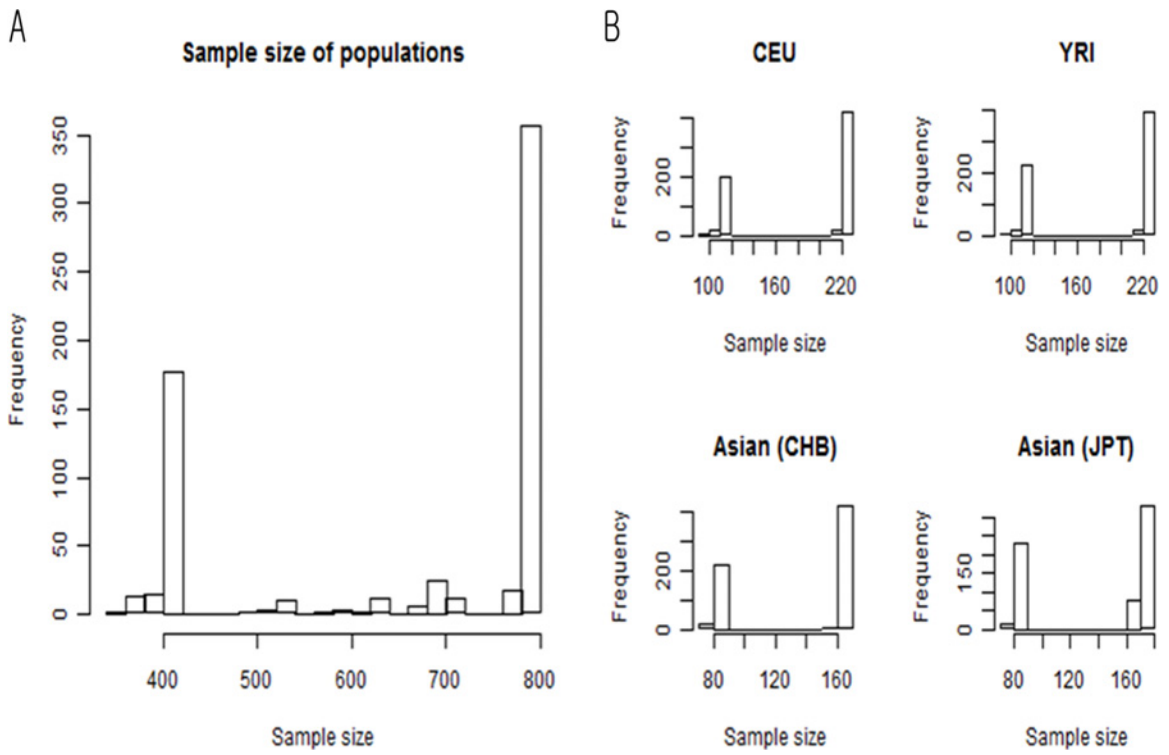

Fig 1. Histogram of sample sizes from 654 drug-related SNPs. A. Total sample sizes of SNPs. **B**. Sample size of each population of SNPs. CHB and JPT are plotted separately according to the format of the original HapMap Data. SNPs with larger sample sizes are included in Phase III, and SNPs with smaller sample sizes are included in Phase II.

doi:10.1371/journal.pone.0119994.g001

(CEU), and East Asian (EA), which consists of the Han Chinese population in Beijing, China (CHB) and the Japanese population in Tokyo, Japan (JPT). Also, we used only founders of CEU and YRI to exclude the related samples. For pharmacological research, we collected 2595 DR SNPs from PharmGKB (http://www.pharmgkb.org) [10,11]. Finally, from these two databases, 654 compatible SNPs among three populations were used for analysis.

2. Comparison of four PD measures using simulation

We selected four PD measures: chi-square test, Weir's F_{st}, ANOVA F-test, and sum of square of d_i from NSCM. Since each phase in HapMap release 27 had different sample sizes, we set the sample size n_i of the subpopulations as a parameter of the simulation as well as the distance d between allele frequencies in order to compare the performance of the four measures. To examine the effect of sample size on these measures, we set n_i as follows:

Scenario I: Increased sample sizes (n_1,n_2,n_3) = (100,100,100), (200,200,200), and (400,400,400).

Scenario II: Unbalanced sample sizes among the subpopulations (n_1,n_2,n_3) = (200,100,100), (100,200,100), and (100,100,200)

For convenience, we assumed equal distance between adjacent alleles and let the distance $d = p_2\text{-}p_1 = p_3\text{-}p_2 (p_1 \leq p_2 \leq p_3)$. We generated p_i with d = 0.1,0.2,0.3 and p_1 was generated under uniform distribution on [0, min(0.5,1-2d)] (Table 2). Here, we used $min(0.5,1\text{-}2d)$ as the maximum of p_1 rather than 1 because of the symmetry in allele frequency. If p_1 is greater than 0.5, then p_2 and p_3 are also greater than 0.5, and we can alternatively identify a set of allele frequencies $\{1\text{-}p_3, 1\text{-}p_2, and\ 1\text{-}p_1\}$ instead of $\{p_1, p_2, p_3\}$.

From these conditions, we generated 200 sets of genotype frequency data from three binomial distributions under Hardy-Weinberg Equilibrium (HWE): binomial distribution(n_i, p_i^2), binomial distribution $(n_i, 2p_iq_i)$ and binomial distribution(n_i, q_i^2). We then calculated the four PD measures. Fig. 2 shows the box plots representing the distributions of the measures from two scenarios. As d increased, the box sizes of the chi-square test, ANOVA F-test, and NSCM increased, while those of F_{st} did not. All measures increased as d increased. As the total sample size increased, the p-values from the chi-square test and ANOVA F-test decreased, while those from F_{st} and SS_d from NSCM did not change significantly (Fig. 2A).

Fig. 2B shows the effect of unbalanced sample size. All measures tended to increase as d increased. All measures showed higher levels of differentiation when n_1 or n_3 was unbalanced $(n_1,n_2,n_3) = (200,100,100)\cdot(100,100,200)$ than when n_2 was unbalanced $(n_1,n_2,n_3) = (100,200,100)$. This was because sample sizes n_1 and n_3 of the subpopulation with extreme values of allele frequencies p_1 and p_3 had large effects on the PD measures.

3. Comparison of the sensitivity and specificity of four PD measures

We simulated additional data from the same sample-size condition n_i as described above. p_1 was generated under uniform distribution on [0, min(0.5,1-2d)]. In these simulations, $d = 0$ indicates the null hypothesis, and other values of d indicate the alternative hypothesis. For a given d and n_i, we generated 100 datasets from the three binomial distributions under HWE. We calculated the specificity (when $d = 0$) and the sensitivity (when $d = 0.05,0.1,\ldots,0.3$) by counting the true negatives and true positives and repeated this step 100 times to calculate the average sensitivity and specificity.

The chi-square test and ANOVA F-test depended on total sample sizes, as indicated by the specificities calculated under the null hypothesis ($d = 0$) for Scenario I (Fig. 3). When the total sample sizes were small, the chi-square test and ANOVA F-test showed high specificities; however, the specificities fell to 92% as the sample size increased. This reflects a general characteristic of test statistics, where the test statistic tends to reject the null hypothesis more when the sample size increases. For Scenario II, all four measures yielded high specificities that were close to one.

In general, the sensitivity increased for all measures as the sample size increased. Still, NSCM consistently yielded the lowest sensitivity. The sensitivities of the chi-square test and ANOVA F-test increased as the total sample size increased. When the sample size was small, F_{st} had the highest sensitivity among the measures (Fig. 4A). When the sample size was moderate or large, the chi-square test and the ANOVA F-test had the highest sensitivities (Fig. 4B, 4C). Note that F_{st} yielded sensitivity and specificity that were robust to sample size, while the other measures did not. Fig. 5 shows the sensitivities from Scenario II. For the same d, specificities were lower when $(n_1,n_2,n_3) = (100,200,100)$ than other situations, similar to the result

Table 2. Examples of data sets.

p_1	d		
	0.1	**0.2**	**0.3**
0	{0, 0.1, 0.2}	{0, 0.2, 0.4}	{0, 0.3, 0.6}
0.1	{0.1, 0.2, 0.3}	{0.1, 0.3, 0.5}	{0.1, 0.4, 0.7}
0.2	{0.2, 0.3, 0.4}	{0.2, 0.4, 0.6}	{0.2, 0.5, 0.8}
0.3	{0.3, 0.4, 0.5}	{0.3, 0.5, 0.7}	{0.3, 0.6, 0.9}
0.4	{0.4, 0.5, 0.6}	{0.4, 0.6, 0.8}	{0.4, 0.7, 1.0}
0.5	{0.5, 0.6, 0.7}	{0.5, 0.7, 0.9}	NA

doi:10.1371/journal.pone.0119994.t002

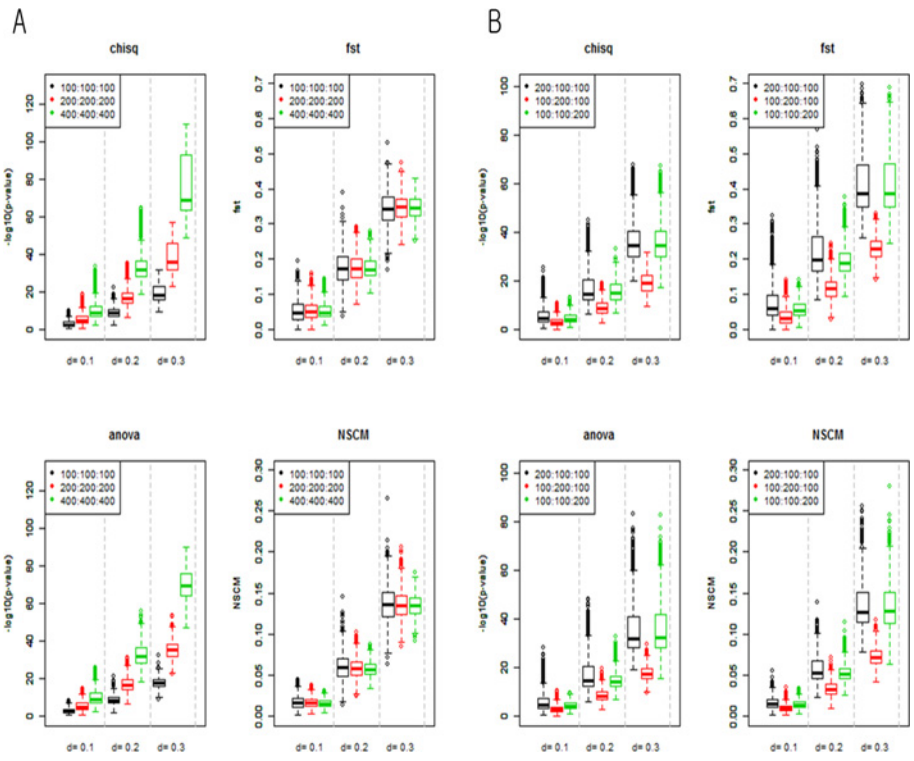

Fig 2. Boxplots representing four measures of simulation data with an increase in *d*. A. Variation of distributions due to increase in sample sizes (Case I). **B.** Variation of distributions due to bias of sample sizes (Case II). For both cases, the *x*-axis denotes the distance *d*, and the *y*-axis and denotes the following measures: $-\log_{10}$ *Pvalue* for chi-square test and ANOVA F-test; Weir and Cockerham's F_{st} estimates for F_{st}; Σd_i^2 for NSCM.

doi:10.1371/journal.pone.0119994.g002

shown in Fig. 2B. NCSM had the lowest sensitivity. The chi-square test and ANOVA F-test had approximately equal sensitivities, while F_{st} had slightly lower sensitivities.

Based on the simulation results, we chose F_{st} for our study for the following reasons. First, the chi-square test and the ANOVA F-test were not appropriate for our data because of their dependence on sample size (Fig. 2A) and low specificity (Fig. 3); this would result in rejecting SNPs without population differences at large sample sizes. NCSM yielded the lowest sensitivity (Fig. 4). Therefore, we chose F_{st} for its high specificity and sensitivity robust to sample size.

4. GO analysis and pathway analysis: comparison of HD and LD gene groups with all genes in DAVID

In order to biologically interpret the HD and LD gene groups, we performed a GO analysis and pathway analysis. Eighteen terms were statistically significant when the HD gene group was analyzed independently, and 48 terms were significant when the LD gene group was analyzed independently. The separate analyses had 25 significant terms in common. Table 3 shows Benjamini-Hochberg's q-values [22] and fold enrichments for GO terms and pathways that were statistically significant in the analysis of the HD gene group only. Table 4 shows the results for the LD gene group. When the LD gene group was used as input for the analysis (Table 4), DR terms in GO categories "drug metabolic process," "drug metabolism," and "metabolism of xenobiotics by cytochrome P450" were significant. The term "drug binding" was

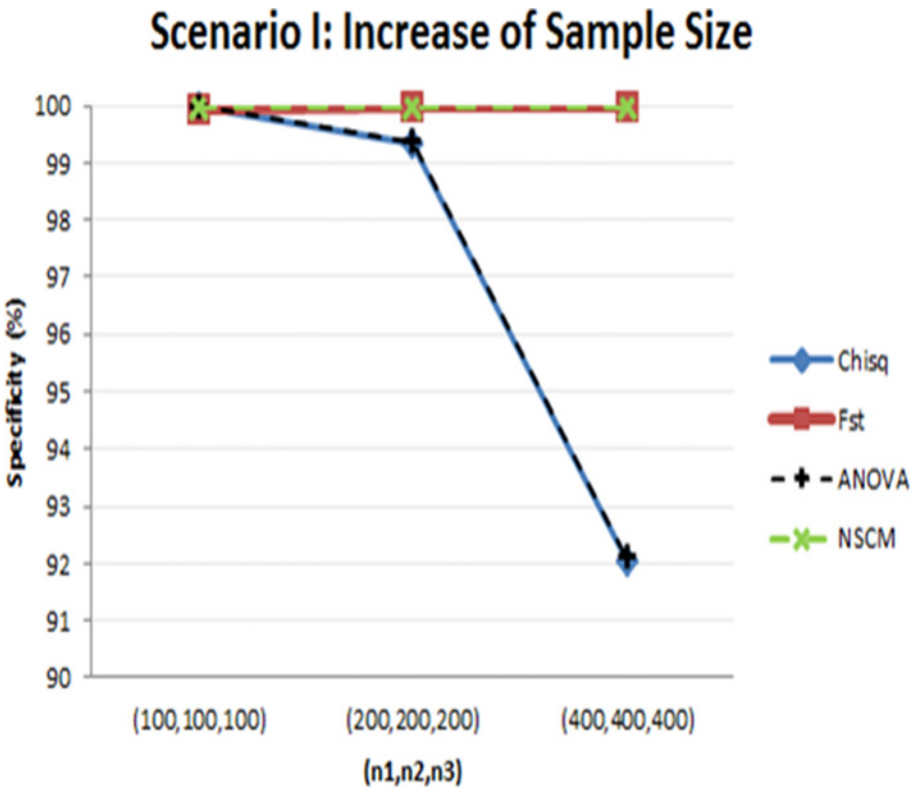

Fig 3. Specificities (%) of each measure from simulation data under H_0:*d* = 0 due to an increase in sample size (Scenario I). The chi-square test and ANOVA F-test are similar, and F_{st} and SS_d from NSCM are nearly identical. Blue line: chi-square test; red line: F_{st}; black dotted line: ANOVA F-test; green dotted line: SS_d from NSCM.

doi:10.1371/journal.pone.0119994.g003

the only significant DR term resulting from analysis of the HD gene group (Table 3). From analysis of the HD gene group, several terms associated with cell communication had significant *p*-values (0.0006, 0.0002, and 0.0142) and high fold enrichments. In particular, the terms, "cell communication" and "cell-cell signaling" had the lowest *p*-values among the terms in Table 3. However, in the LD gene group, only one significant term was related to cell communication ("regulation of cell communication"), which had a q-value (0.0273) and fold enrichment (2.41).

Seventeen genes were associated with the three terms related to cell communication ("cell-cell signaling," "cell communication," and "drug binding"). Among these 17 genes, we extracted five that contained SNPs with F_{st} values greater than 0.5: *STX4*, *PPARD*, *DCK*, *GRIK4*, *and DRD3* contained SNPs rs10871454, rs6922548, rs3775289, rs1954787, and rs167771 with F_{st} values of 0.682, 0.620, 0.573, 0.531, and 0.510, respectively.

Syntaxin 4 (STX4) is a component of the SNARE complex, which mediates docking of cellular transport vesicles. In a GWAS, rs10871454 in *STX4* accounted for over 25% of the variance in log-transformed stabilized warfarin dose and was in perfect linkage disequilibrium with rs9923231 [33].

PPARs are nuclear hormone receptors that bind peroxisome proliferators and control the size and number of peroxisomes produced by cells. In particular, PPARδ is a receptor that binds peroxisome proliferators such as hypo-lipidemic drugs and fatty acids [34]. The SNP

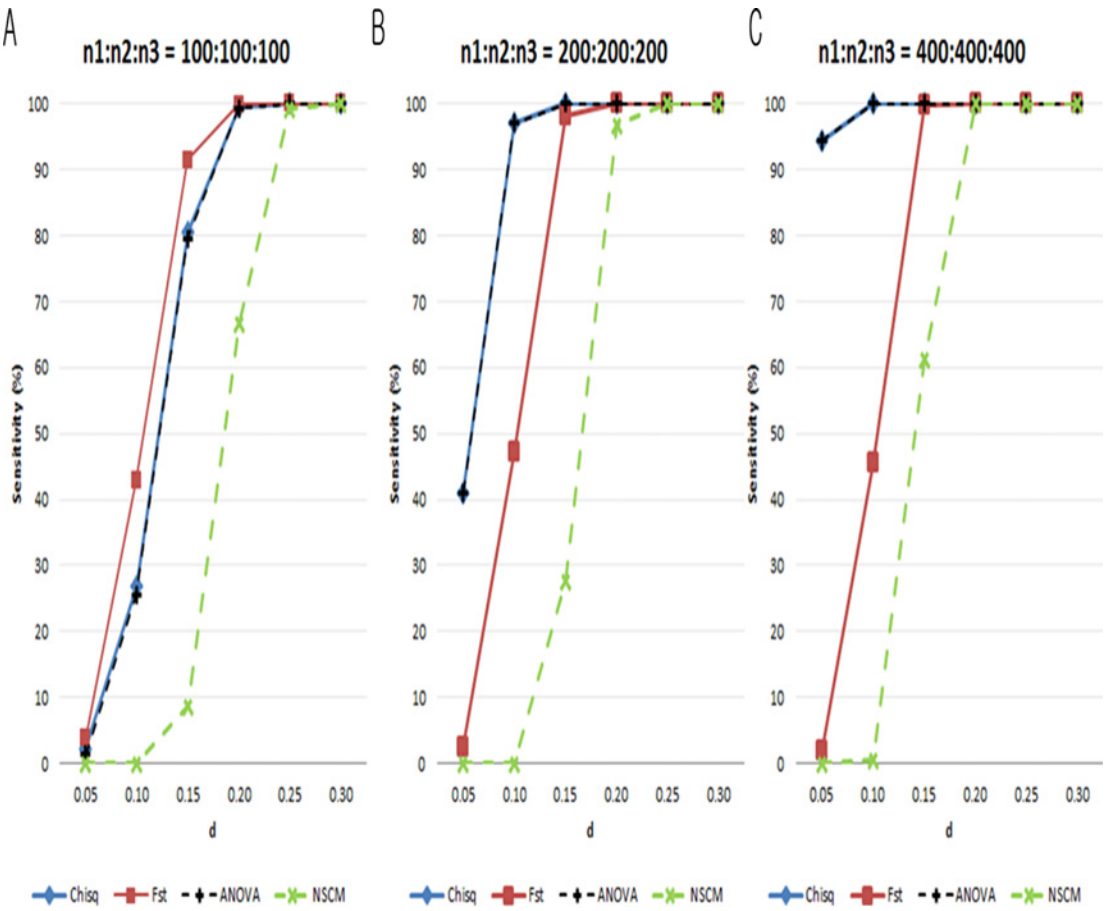

Fig 4. Sensitivities (%) of each measure from simulation data under H_0:*d* = 0.05,0.1. . .,0.3 due to an increase in sample size (Scenario I). **A**. (n_1,n_2,n_3) = (100,100,100). **B**. (n_1,n_2,n_3) = (200,200,200). **C**. (n_1,n_2,n_3) = (100,100,100). Blue line: chi-square test; red line: F_{st}; black dotted line: ANOVA F-test; green dotted line: SS_d from NSCM.

doi:10.1371/journal.pone.0119994.g004

rs6922548 in *PPARδ* is also associated with positive clinical response to docetaxel and thalidomide [35].

DCK is involved in the gemcitabine and lamivudine pathways [36,37]. It also participates in the phosphorylation of cytosolic nucleosides by deoxycytidine kinase and pyrimidine salvage reactions. Fukunaga *et al.* examined variants that were identified from a screen of 13 genes in the gemcitabine metabolic pathway [38] and found that the C allele of SNP rs3775289 was not present among Europeans or Africans in their study. However, in the International HapMap Data, the allelic frequencies of the C allele are 0.929 in Europeans and 0.279 in Africans.

Paddock *et al.* implemented an association study based on the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) cohort and found that rs1954787 in the *GRIK4* gene, which encodes the kainic acid-type glutamate receptor KA1, was associated with response to the antidepressant citalopram [39]. Accordingly, they suggested that the glutamate system plays an important role in modulating response to selective serotonin reuptake inhibitors (SSRIs). In addition, Pickard *et al.* showed that variation in *GRIK4* was significantly associated with both an increased risk of schizophrenia and a decreased risk of bipolar disorder [40]. Furthermore, the G variant of SNP rs167771 in *DRD3* was associated with an increased risk of extrapyramidal symptoms (EPS) in psychiatric patients receiving risperidone [41].

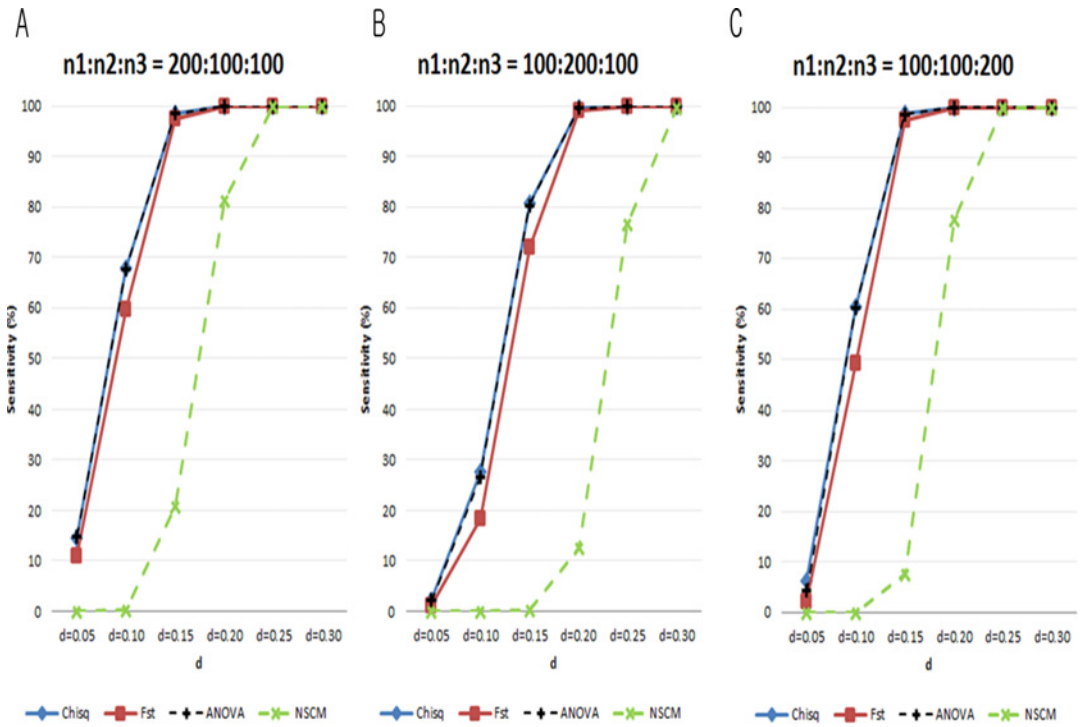

Fig 5. Sensitivities (%) of each measure from simulation data under H_0:*d*=0.05,0.1. . .,0.3 due to bias in sample size (Scenario II). A. (n_1,n_2,n_3) = (200,100,100). **B**. (n_1,n_2,n_3) = (100,200,100). **C**. (n_1,n_2,n_3) = (100,100,200). Blue line: chi-square test; red line: F_{st}; black dotted line: ANOVA F-test; green dotted line: SS_d from NSCM.

doi:10.1371/journal.pone.0119994.g005

Several terms related to reproduction were identified in the analysis of the HD gene group, but none were identified in the analysis of the LD gene group. For example, Table 3 reports the terms "sex differentiation," "development of primary sexual characteristics," and "reproductive developmental process." Similarly, Wu and Zhang reported that reproduction-associated processes (*e.g.*, "sperm motility," "spermatid development," and "gamete generation") had higher levels of PD [1]. The criteria for HD genes in the present study were more robust than those for genes with higher levels of PD used by Wu and Zhang. Nevertheless, Table 3 reports several

Table 3. Q-values and fold enrichments of significant terms in HD group.

Terms	FE	BH's q	Terms	FE	BH's q
BP: Behavioral response to nicotine	96.16	0.0201	BP: Secretion	5.24	0.0466
MF: Drug binding	16.51	0.0142	BP: Regulation of cell differentiation	4.56	0.0157
BP: Adult behavior	15.65	0.0075	BP: Cell Communication	4.53	0.0002
BP: Regulation of multicellular organism growth	14.71	0.0465	BP: Cell development	3.48	0.0464
CC: Dendrite	9.63	0.0027	BP: Regulation of developmental process	3.34	0.0276
BP: Sex differentiation	8.92	0.0210	BP: Neurological system process	3.15	0.0060
BP: Development of primary sexual characteristics	8.83	0.0468	BP: Anatomical structure morphogenesis	2.81	0.0197
BP: Reproductive developmental process	5.99	0.0307	BP: Cell differentiation	2.47	0.0214
BP: Cell-cell signaling	5.24	0.0006	BP: System development	2.21	0.0136

FE: Fold enrichment

BH's q: Benjamini-Hochberg's q-value

doi:10.1371/journal.pone.0119994.t003

Table 4. Q-values and fold enrichments of significant terms in LD group.

Terms	FE	BH's q	Terms	FE	BH's q
BP: Negative Regulation Of Amine Transport	58.70	0.0170	BP: Regulation Of Response To Stress	5.71	0.0017
*BP: Drug Metabolic Process	48.92	0.0003	BP: Positive Regulation Of Transport	5.62	0.0103
BP: Negative Regulation Of Glucose Transport	46.96	0.0253	CC: Membrane fraction	5.33	4.4E-12
BP: Negative Regulation Of Organic Acid Transport	46.96	0.0253	CC: Insoluble fraction	5.31	3.0E-12
BP: Multicellular Organismal Water Homeostasis	42.69	0.0281	BP: Heterocycle Metabolic Process	5.16	0.0017
BP: Body Fluid Secretion	28.99	0.0011	BP: Positive Regulation Of Multicellular Organismal Process	5.13	0.0144
BP: Negative Regulation Of Ion Transport	27.23	0.0082	BP: Regulation Of Anatomical Structure Morphogenesis	5.00	0.0329
MF: Oxygen Binding	26.82	0.0000	BP: Cellular Chemical Homeostasis	4.94	0.0011
BP: Renal System Process	22.36	0.0126	CC: Cell projection part	4.91	0.0173
CC: Presynaptic membrane	19.81	0.0167	CC: Microsome	4.85	0.0173
BP: Regulation Of Tube Size	17.72	0.0012	CC: Vesicular fraction	4.71	0.0190
KEGG: Linoleic acid metabolism	14.24	0.0306	KEGG: Calcium signaling pathway	4.53	0.0217
*KEGG: Drug metabolism: other enzymes	13.91	0.0024	CC: Cell fraction	4.51	3.1E-12
KEGG: Arachidonic acid metabolism	12.46	0.0013	BP: Anatomical Structure Formation Involved In Morphogenesis	4.00	0.0257
MF: Tetrapyrrole Binding	11.49	0.0000	BP: Transmembrane Transport	3.85	0.0018
KEGG: Retinol metabolism	11.08	0.0048	MF: Oxidoreductase Activity	3.83	0.0002
*KEGG: Metabolism of xenobiotics by cytochrome P450	9.97	0.0060	BP: Regulation Of Response To Stimulus	3.70	0.0124
*KEGG: Drug metabolism: cytochrome p450	9.65	0.0056	BP: Ion Transport	3.67	0.0005
BP: Negative Regulation Of Multicellular Organismal Process	8.59	0.0006	BP: Positive Regulation Of Molecular Function	3.47	0.0065
BP: Negative Regulation Of Transport	8.12	0.0051	BP: Homeostatic Process	3.13	0.0055
BP: Regulation Of Response To External Stimulus	7.88	0.0019	BP: Regulation Of Multicellular Organismal Process	3.01	0.0018
BP: Regulation Of Body Fluid Levels	6.66	0.0281	BP: Regulation Of Catalytic Activity	2.78	0.0133
BP: Muscle System Process	6.52	0.0128	CC: Endomembrane system	2.75	0.0169
BP: Angiogenesis	6.35	0.0317	BP: Regulation Of Cell Communication	2.41	0.0273

FE: Fold enrichment

BH's q: Benjamini-Hochberg's q-value

doi:10.1371/journal.pone.0119994.t004

terms also identified by Wu and Zhang [1]. For example, terms related to the nervous system ("dendrite" and "neurological system process"), development ("anatomical structure morphogenesis," "regulation of developmental process," "system development," and "cell development"), stress response, homeostasis, growth ("regulation of multicellular organism growth"), secretion ("secretion"), and metabolism had high levels of PD.

The term "behavioral response to nicotine" had the highest fold enrichment of 96.16 (Table 3). This was likely due to the large number of nicotine-related SNPs in our dataset. Specifically, 97 of 654 SNPs were associated with nicotine according to the annotation provided by PharmGKB. The term "adult behavior," which had a high fold enrichment (15.65) and low *p*-value (0.0075), is an ancestor term of "behavioral response to nicotine."

Table 5. Q-values and fold enrichments of hypergeometric test between HD group and drug-related (DR) genes.

Terms	HD vs. others		DR vs. others		HD vs DR	
	FE	BH's q	FE	BH's q	FE	BH's q
BP: Sex Differentiation	8.92	0.0210	2.52	0.2015	3.54	0.0158
BP: Development Of Primary Sexual Characteristics	8.83	0.0468	2.57	0.2558	3.44	0.0283
BP: Reproductive structure development	8.90	0.0504	2.58	0.2531	3.44	0.0283
BP: Gonad development	9.74	0.1139	2.81	0.3622	3.47	0.0527
BP: Development Of Primary Sexual Characteristics	8.83	0.0468	2.57	0.2558	3.47	0.0527
KEGG: Wnt signaling pathway	3.45	0.6823	1.32	0.8715	2.62	0.0556

FE: Fold enrichment

BH's q: Benjamini-Hochberg's q-value

doi:10.1371/journal.pone.0119994.t005

Two terms were associated with differentiation in Table 3. "Regulation of cell differentiation" and "cell differentiation" showed fold enrichments of 4.56 and 2.47 and q-values of 0.0157 and 0.0214, respectively.

5. GO analysis and pathway analysis: comparison to DR genes

We performed GO analysis and pathway analysis in order to biologically interpret the relationship between the HD gene group and DR genes. The results are summarized as q-values using Benjamini-Hochberg's method [41] and fold enrichments of the GO terms and pathways. A Benjamini-Hochberg's q-value less than 0.05 indicates that the HD group contained significantly more genes in the term as compared to randomly selected DR genes.

Table 5 shows the results from the GO analysis and pathway analysis comparing the HD gene group and DR genes. The resulting terms included those related to reproduction that were included in Table 3: "sex differentiation," "development of primary sexual characteristics," "reproductive structure development," "gonad development," and "development of primary sexual characteristics."

Since Wu and Zhang [1] did not conduct a pathway analysis, the "Wnt signaling pathway" was not directly identified. However, their GO analysis identified the term "Wnt receptor signaling pathway through beta-catenin" as statistically significant in the HD gene group [1]. The Wnt signaling pathway is important in pharmacogenetics, because it is strongly associated with cancer [42,43]. Further studies are warranted to identify drugs that inhibit the Wnt signaling pathway, because inhibition of aberrant Wnt signaling in cancer cell lines inhibits their growth [44].

Discussion

PD is important for understanding differences in drug responses among populations. However, PD often refers to the distance between two different subpopulations; therefore, several studies have investigated approaches for averaging the PD of each SNP. For instance, the impact of SNP ascertainment on estimating the distance between subpopulations has already been reported [45]. In contrast, the present study identified population-specific pharmacogenomics variants. We did not focus on identifying average distances using all SNPs; rather, we used each SNP to identify population-specific pharmacogenomics variants. As a result, our results described the impact of sample ascertainment on different measures of PD for each SNP. In addition, the present study investigated PD of genes in the PharmGKB database, while several previous studies have focused on genes related to individual drugs [5]. This approach enabled

us to more systematically study PD of DR genes by considering all reported DR genes from PharmGKB.

In our comparison study, F_{st} showed high specificity and sensitivity robust to the different sample sizes of HapMap release 27. After calculating F_{st} from the allele frequency data of each SNP, we defined HD and LD gene groups. Then, we performed GO analysis and pathway analysis to describe the biological characteristics of the HD gene group. We compared the HD gene group to two different backgrounds: all genes in DAVID and DR genes in the PharmGKB database.

The GO and pathway analyses identified two terms related to cell communication ("cell-cell signaling" and "cell communication"), which had the lowest *p*-values (0.0006 and 0.0002, respectively). In addition, the term "drug binding," which was related to cell communication, was also considered to be meaningful due to its high fold enrichment (16.51) despite its moderate *p*-value (0.0142). Thus, these results suggest that the HD gene group from PharmGKB is highly associated with cell communication. Since drug binding is associated with the cell membrane, similar to processes related to both cell-cell signaling and cell communication, the simultaneous identification of these GO terms is convincing. In addition, this finding suggests that the cellular location of gene products affects PD. It is possible that, the outer surface of the cell membrane is initially affected by mutagens, because it is closest to the extracellular environment.

In addition, we examined genes containing SNPs with high F_{st} values (above 0.5) among cell-communication-related terms, such as *STX4*, *PPARD*, *DCK*, *GRIK4*, and *DRD3*. Specifically, SNPs rs10871454, rs6922548, rs3775289, rs1954787, and rs167771 had F_{st} values of 0.682, 0.620, 0.573, 0.531, and 0.510, respectively. Further biological studies of these genes will help elucidate their roles in pharmacogenetics.

Unlike other GO analyses, we employed DR genes from PharmGKB and performed an additional analysis by using them as a background for the GO analysis. This strategy of changing the background of the GO analysis from all genes to DR genes represents a novel method. Therefore, our approach has the advantage of providing distinct information about genes of interest by altering the background of analysis.

There are several similarities and differences between the PD study of Wu and Zhang [1] and the present study. Both studies determined PD using F_{st} and investigated characteristics of genes with a high level of differentiation by GO analysis. Thus, the studies identified several similar terms such as those related to replication, development, and metabolism. In addition to the HD gene group, our study performed GO analysis of the LD gene group and compared the results, which identified distinct characteristics of each group. We also conducted GO and pathway analyses comparing the HD gene group to DR genes and identified two meaningful terms, "drug binding" and "Wnt signaling pathway," which were not identified by Wu and Zhang.

In conclusion, the present study describes an approach for assessing PD associated with multiple drugs using a database. Therefore, the integrated approach may identify valid genetic features different from the background gene list. We validated results from other systematic analyses. Moreover, our approach allows the possibility of improving the results. DR genes that are unknown or newly reported were not included in the present study. Thus, our approach may be limited in its ability to interpret the population-specific difference in drug response or efficacy caused by genetic divergence. However, this method remains convincing, because our statistical analyses revealed high specificity and sensitivity robust to sample size. Furthermore, we obtained significant differences from other DR genes in the PharmGKB database, and our approach thus represents a systematic method for identifying valid population-specific pharmacogenomics variants.

Author Contributions

Conceived and designed the experiments: KIK IWK JMO TP. Performed the experiments: BY KIK IWK JMO EA TP. Analyzed the data: BY KIK IWK JMO EA TP. Wrote the paper: BY KIK IWK JMO EA TP.

References

1. Wu DD, Zhang YP (2011) Different level of population differentiation among human genes. Bmc Evolutionary Biology 11.
2. Myles S, Davison D, Barrett J, Stoneking M, Timpson N (2008) Worldwide population differentiation at disease-associated SNPs. Bmc Medical Genomics 1.
3. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L (2008) Natural selection has driven population differentiation in modern humans. Nature Genetics 40: 340–345. doi: 10.1038/ng.78 PMID: 18246066
4. Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG, et al. (2001) Population genetic structure of variable drug response. Nature Genetics 29: 265–269. PMID: 11685208
5. Gage BF, Lesko LJ (2008) Pharmacogenetics of warfarin: regulatory, scientific, and clinical issues. Journal of Thrombosis and Thrombolysis 25: 45–51. PMID: 17906972
6. Takahashi H, Wilkinson GR, Caraco Y, Muszkat M, Kim RB, Kashima T, et al. (2003) Population differences in S-warfarin metabolism between CYP2C9 genotype-matched Caucasian and Japanese patients. Clinical Pharmacology & Therapeutics 73: 253–263.
7. Pavani A, Naushad SM, Rupasree Y, Kumar TR, Malempati AR, Pinjala RK, et al. (2012) Optimization of warfarin dose by population-specific pharmacogenomic algorithm. Pharmacogenomics Journal 12: 306–311. doi: 10.1038/tpj.2011.4 PMID: 21358752
8. Huang RS, Duan SW, Bleibel WK, Kistner EO, Zhang W, Clark TA, et al. (2007) A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. Proceedings of the National Academy of Sciences of the United States of America 104: 9758–9763. PMID: 17537913
9. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851–861. PMID: 17943122
10. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. (2012) Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther 92: 414–417. doi: 10.1038/clpt.2012.96 PMID: 22992668
11. Hewett M, Oliver DE, Rubin DL, Easton KL, Stuart JM, Altman RB, et al. (2002) PharmGKB: The Pharmacogenetics Knowledge Base. Nucleic Acids Research 30: 163–165. PMID: 11752281
12. Joubert BR, North KE, Wang YF, Mwapasa V, Franceschini N, Meshnick SR, et al. (2010) Comparison of genome-wide variation between Malawians and African ancestry HapMap populations. Journal of Human Genetics 55: 366–374. doi: 10.1038/jhg.2010.41 PMID: 20485449
13. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Research 12: 1805–1814. PMID: 12466284
14. Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population-Structure. Evolution 38: 1358–1370.
15. Weir BS, Hill WG (2002) Estimating F-statistics. Annual Review of Genetics 36: 721–750. PMID: 12359738
16. Voight BF, Kudaravalli S, Wen XQ, Pritchard JK (2006) A map of recent positive selection in the human genome. Plos Biology 4: 446–458.
17. Casto AM, Feldman MW (2011) Genome-wide association study SNPs in the human genome diversity project populations: does selection affect unlinked SNPs with shared trait associations? PLoS Genet 7: e1001266. doi: 10.1371/journal.pgen.1001266 PMID: 21253569
18. Park J, Hwang S, Lee YS, Kim SC, Lee D (2007) SNP@Ethnos: a database of ethnically variant single-nucleotide polymorphisms. Nucleic Acids Res 35: D711–D715. PMID: 17135185
19. Tibshirani R, Hastie T, Narasimhan B, Chu G (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proceedings of the National Academy of Sciences of the United States of America 99: 6567–6572. PMID: 12011421
20. Han K, Kim K, Park T (2010) Unbalanced sample size effect on the genome-wide population differentiation studies; pp. 347–352.

21. Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R (2014) On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. Molecular Biology and Evolution.
22. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 57: 289–300.
23. Li Y, Zhu YM, Liu Y, Shu YJ, Meng FJ, Lu YM, et al. (2008) Genome-wide identification of osmotic stress response gene in Arabidopsis thaliana. Genomics 92: 488–493. doi: 10.1016/j.ygeno.2008.08.011 PMID: 18804526
24. Wright S (1978) Evolution and the genetics of populations. Chicago: University of Chicago Press.
25. Cardoso MA, Provan J, Powell W, Ferreira PCG, De Oliveira DE (1998) High genetic differentiation among remnant populations of the endangered Caesalpinia echinata Lam. (Leguminosae-Caesalpinioideae). Molecular Ecology 7: 601–608.
26. Strauss SH, Hong YP, Hipkins VD (1993) High-Levels of Population Differentiation for Mitochondrial-DNA Haplotypes in Pinus-Radiata, Muricata, and Attenuata. Theoretical and Applied Genetics 86: 605–611. doi: 10.1007/BF00838716 PMID: 24193710
27. Duan S, Zhang W, Cox NJ, Dolan ME (2008) FstSNP-HapMap3: a database of SNPs with high population differentiation for HapMap3. Bioinformation 3: 139–141. PMID: 19238253
28. Sherman BT, Huang DW, Tan QN, Guo YJ, Bour S, Liu D, et al. (2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. Bmc Bioinformatics 8.
29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. (2000) Gene Ontology: tool for the unification of biology. Nature Genetics 25: 25–29. PMID: 10802651
30. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28: 27–30. PMID: 10592173
31. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467: 52–58. doi: 10.1038/nature09298 PMID: 20811451
32. Kim IW, Kim KI, Chang HJ, Yeon B, Bang SJ, Park TS, et al. (2012) Ethnic variability in the allelic distribution of pharmacogenes between Korean and other populations. Pharmacogenetics and Genomics 22: 829–836. doi: 10.1097/FPC.0b013e328358dd70 PMID: 22955668
33. Teichert M, Eijgelsheim M, Uitterlinden AG, Buhre PN, Hofman A, De-Smet PAGM, et al. (2011) Dependency of phenprocoumon dosage on polymorphisms in the VKORC1, CYP2C9, and CYP4F2 genes. Pharmacogenetics and Genomics 21: 26–34. doi: 10.1097/FPC.0b013e32834154fb PMID: 21063236
34. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. (2014) DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 42: 1091–1097.
35. Deeken JF, Cormier T, Price DK, Sissung TM, Steinberg SM, Tran K, et al. (2010) A pharmacogenetic study of docetaxel and thalidomide in patients with castration-resistant prostate cancer using the DMET genotyping platform. Pharmacogenomics Journal 10: 191–199. doi: 10.1038/tpj.2009.57 PMID: 20038957
36. Kroep JR, Loves WJP, van der Wilt CL, Alvarez E, Talianidis I, Boven E, et al. (2002) Pretreatment Deoxycytidine Kinase Levels Predict in Vivo Gemcitabine Sensitivity 1 Supported by Eli Lilly & Co, International and The Netherlands.1. Molecular Cancer Therapeutics 1: 371–376. PMID: 12477049
37. Kewn S, Hoggard P, Sales SD, Johnson MA, Back DJ (2001) The intracellular activation of lamivudine (3TC) and determination of 2'-deoxycytidine-5'-triphosphate (dCTP) pools in the presence and absence of various drugs in HepG2 cells. British Journal of Clinical Pharmacology 50: 597–604.
38. Fukunaga AK, Marsh S, Murry DJ, Hurley TD, McLeod HL (2004) Identification and analysis of single-nucleotide polymorphisms in the gemcitabine pharmacologic pathway. Pharmacogenomics Journal 4: 307–314. PMID: 15224082
39. Paddock S, Laje G, Charney D, Rush AJ, Wilson AF, Sorant AJM, et al. (2007) Association of GRIK4 with outcome of antidepressant treatment in the STAR*D cohort. American Journal of Psychiatry 164: 1181–1188. PMID: 17671280
40. Pickard BS, Knight HM, Hamilton RS, Soares DC, Walker R, Boyd JFK, et al. (2008) A common variant in the 3' UTR of the GRIK4 glutamate receptor gene affects transcript abundance and protects against bipolar disorder. Proceedings of the National Academy of Sciences of the United States of America 105: 14940–14945. doi: 10.1073/pnas.0800643105 PMID: 18824690
41. Gasso P, Mas S, Bernardo M, Alvarez S, Parellada E, Lafuente A. (2009) A common variant in DRD3 gene is associated with risperidone-induced extrapyramidal symptoms. Pharmacogenomics Journal 9: 404–410. doi: 10.1038/tpj.2009.26 PMID: 19506579
42. Garber K (2009) Drugging the Wnt Pathway: Problems And Progress. Journal of the National Cancer Institute 101: 548–550. doi: 10.1093/jnci/djp084 PMID: 19351922

43. Takahashi-Yanaga F, Sasaguri T (2007) The Wnt/beta-catenin signaling pathway as a target in drug discovery. Journal of Pharmacological Sciences 104: 293–302. PMID: 17721040

44. Barker N, Clevers H (2006) Mining the Wnt pathway for cancer therapeutics. Nature Reviews Drug Discovery 5: 997–1014. PMID: 17139285

45. Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting FST: The impact of rare variants. Genome Research 23: 1514–1521. doi: 10.1101/gr.154831.113 PMID: 23861382