

Prediction and visualization of CYP2D6 genotype-based phenotype using clustering algorithms

Eun-Young Kim^{1,2}, Sang-Goo Shin³ and Jae-Gook Shin^{1,2*}

¹Department of Clinical Pharmacology, Inje University College of Medicine, Busan Paik Hospital, Busan 47392, Republic of Korea,

²Department of Pharmacology and Pharmacogenomics Research Center, Inje University College of Medicine, Busan 47392, Republic of Korea, ³Department of Clinical Pharmacology and Therapeutics, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

*Correspondence: J. G. Shin; Tel: +82-51-890-6709, fax: +82-51-893-1232, E-mail: phshinjg@gmail.com



Received 2 Aug 2017

Revised 5 Sep 2017

Accepted 5 Sep 2017

Keywords

CYP2D6,
clustering analysis,
dextromethorphan,
genotype,
phenotype

pISSN: 2289-0882

eISSN: 2383-5427

This study focused on the role of cytochrome P450 2D6 (CYP2D6) genotypes to predict phenotypes in the metabolism of dextromethorphan. *CYP2D6* genotypes and metabolic ratios (MRs) of dextromethorphan were determined in 201 Koreans. Unsupervised clustering algorithms, hierarchical and k-means clustering analysis, and color visualizations of CYP2D6 activity were performed on a subset of 130 subjects. A total of 23 different genotypes were identified, five of which were observed in one subject. Phenotype classifications were based on the means, medians, and standard deviations of the log MR values for each genotype. Color visualization was used to display the mean and median of each genotype as different color intensities. Cutoff values were determined using receiver operating characteristic curves from the k-means analysis, and the data were validated in the remaining subset of 71 subjects. Using the two highest silhouette values, the selected numbers of clusters were three (the best) and four. The findings from the two clustering algorithms were similar to those of other studies, classifying *5/*5 as a lowest activity group and genotypes containing duplicated alleles (i.e., *CYP2D6*1/*2N*) as a highest activity group. The validation of the k-means clustering results with data from the 71 subjects revealed relatively high concordance rates: 92.8% and 73.9% in three and four clusters, respectively. Additionally, color visualization allowed for rapid interpretation of results. Although the clustering approach to predict CYP2D6 phenotype from *CYP2D6* genotype is not fully complete, it provides general information about the genotype to phenotype relationship, including rare genotypes with only one subject.

Introduction

Genetic polymorphism in drug-metabolizing enzymes can cause inter-individual differences in drug response. Because of the pharmacokinetic and pharmacodynamic diversity of these drugs, the same dose can result in a variety of clinical conse-

quences in different patients, such as adverse drug reactions or therapeutic failure.[1,2] Cytochrome P450 (CYP) enzymes play an important role in the metabolism of endogenous substrates and xenobiotics.[3,4] More than 50 CYP enzymes have been identified so far.[5]

The CYP2D6 enzyme is involved in the metabolism of more than 100 drugs, including antipsychotics, antiarrhythmics, and analgesics, which account for approximately 15–25% of all clinically used drugs.[6,7] *CYP2D6* is highly polymorphic and more than 100 allelic variants have been reported.[8] A range of enzyme activity results from the broad spectrum of genetic variation, point mutation, defect or insertion, genetic rearrangement, and deletion or duplication of the whole gene.

Copyright © 2017 Translational and Clinical Pharmacology
© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>).
© This paper meets the requirement of KS X ISO 9706, ISO 9706-1994 and ANSI/NISO Z.39.48-1992 (Permanence of Paper).

Reviewer

This article was reviewed by peer experts who are not TCP editors.

[7] These *CYP2D6* genetic polymorphisms are considered as distinct ethnic differences in drug metabolism depending on each gene variation. Enzyme activity, defined as the degree of metabolism, can be classified into four groups: ultra-rapid metabolizers (UM), extensive metabolizers (EM), intermediate metabolizers (IM), and poor metabolizers (PM).[7,9] The Clinical Pharmacogenetics Implementation Consortium recommends that normal metabolizer (NM) be the standard term in place of EM.[10] Nortriptyline, a tricyclic antidepressant primarily metabolized by *CYP2D6*, requires < 10–20 mg for PMs, and up to 500 mg for UMs, to achieve the desired therapeutic effect.[1] This suggests that treatment outcomes are dependent on drug metabolism capacity, as determined by the genetic predisposition of *CYP2D6*. At the same dose, PM patients may suffer from toxic effects, while UM patients may not be affected. Knowledge of a patient's *CYP2D6* genotype prior to drug therapy allows prediction of enzyme activity, which can be useful for optimization of individual therapies. Many clinically important drugs metabolized by *CYP2D6* have narrow therapeutic dose ranges especially in PMs, so genotype information for *CYP2D6* has the potential to be beneficial to patients.

In this study, *CYP2D6* genotypes were classified based on *CYP2D6* metabolic activity to predict genotype-based phenotypes in Korean subjects. The results of the study can be used for direct individualized therapy based on *CYP2D6* genotypes. The relationship between genotype and phenotype was assessed using machine learning algorithms. Use of color visualizations allowed for easier identification of phenotype based on *CYP2D6* genotype.

Methods

CYP2D6 genotype and phenotype data

CYP2D6 phenotyping data performed in 201 healthy Korean subjects using dextromethorphan (DM) as the *CYP2D6* substrate were used. The Institutional Review Board of the Inje University Busan Paik Hospital (Busan, Republic of Korea) approved the study protocol.(06-087) The genotype and phenotype data retrieved from the database of the Pharmacogenomics Research Center (PGRC), Inje University (Busan, Republic of Korea),[11,12] were used for this study. The method of phenotyping with urine dextromethorphan and its metabolite is described in the previous study.[11] Genotypes were identified using the multiplex single-base extension method to select 12 *CYP2D6* alleles (*1, *2, *5, *10, *14, *18, *21, *41, *49, *52, *60, and a duplication of *CYP2D6*).[12] *CYP2D6* activity was measured as the log-transformed urinary metabolic ratio (MR) of DM, log₁₀ (DM/Dextrophan (DX)).[11]

CYP2D6 genotype to phenotype classification analysis

Cluster analysis was performed to classify *CYP2D6* genotypes according to phenotype (based on the mean, median, and standard deviation (SD) of log-transformed MR) using data from a subset of 130 subjects (clustering set). Two algorithms were

applied for the cluster analysis: hierarchical clustering and k-means clustering. In hierarchical clustering, ward linkage was selected as the agglomeration procedure and Euclidean distance was used to calculate similarities between individuals. The number of likely groups was determined by visual observation of a dendrogram. In k-means clustering analysis, the number of phenotype clusters was assessed by calculating the average of the silhouette values obtained by 20 different initial values. K-means clustering with squared Euclidean distance was performed with the two numbers of clusters with the highest silhouette values. The calculated silhouette value ranges from -1 to 1. A value ≥ 0 indicates that the value should be classified into the group, while a value of -1 indicates the value is not related to the group. A value close to 1 indicates a well-separated group.

Color visualization and validation of phenotype classifications

Hierarchical and k-means clustering were plotted on a map using color to indicate phenotype (i.e., the level of drug metabolism activity) for each genotype. A dendrogram and silhouette values were also visualized in hierarchical and k-means clustering, respectively. Higher *CYP2D6* activity levels were represented by the color blue, while higher logarithmic MR values (lower enzyme activity) were represented by red.

To verify the phenotype classification determined using the k-means analysis in the clustering set (130 subjects), the results were applied to the validation set (71 subjects who were recruited separately). The following procedure was used:

- 1) The cutoff value between two clusters in proximal activity groups was determined using receiver operating characteristic (ROC) curves as the log MR value of the individuals belonging to each cluster in the clustering set (130 subjects). The lower enzyme activity of the two clusters determined the cutoff value. The cutoff value was nearest to the upper left corner of the ROC curve with the most satisfactory sensitivity and specificity.
- 2) The cutoff value was used to determine the concordance rate in the validation set (71 subjects), that is, whether each individual was within a phenotype cluster including his/her genotype.

Cluster analysis, determination of cutoff values using ROC curves, and color visualizations were performed using MATLAB 7.1 (MathWorks Inc., Natick, MA, USA).

Results

Subject characteristics (age: 25 ± 3.6 years) have been previously described.[11] There were 23 different genotypes in the clustering set (130 subjects), five of which were observed in one subject only. The most common genotypes were *CYP2D6**1/*10B, *10B/*10B, and *1/*1, occurring at frequencies of 20.0%, 19.2%, and 16.2%, respectively. Other genotypes occurred at significantly lower frequencies. Sixteen genotypes were identified in the validation set of 71 subjects, the most common of which

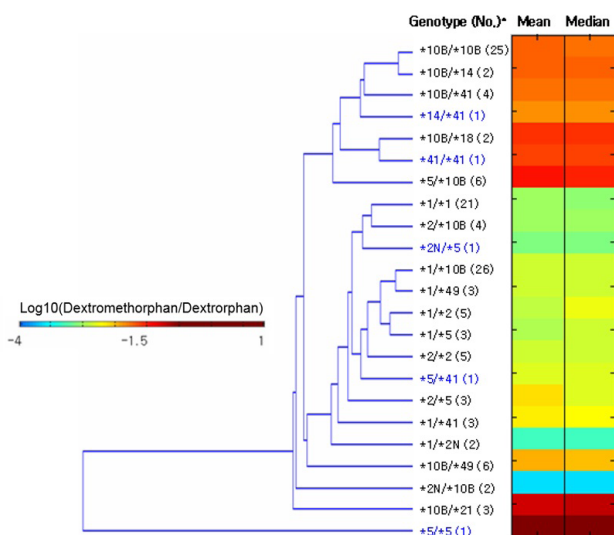


Figure 1. Color visualization of log-transformed urinary metabolic ratios (MR, dextromethorphan/dextrorphan) in hierarchical clustering (n=130). Each row indicates a *CYP2D6* genotype. The mean and the median log MR values are colored based on intensity (cold to warm color). Genotypes indicated in blue font are based on data from a single subject. * No., number of subjects.

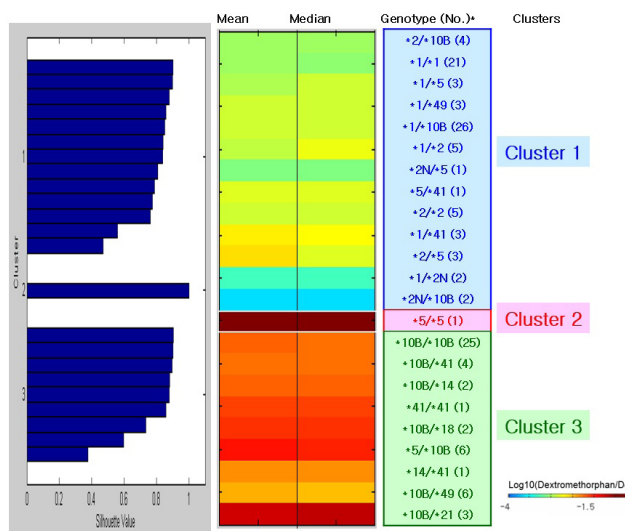


Figure 2. Color visualization and silhouette values of log-transformed urinary metabolic ratios (MR, dextromethorphan/dextrorphan) using the k-means cluster analysis based on three clusters (n=130). Each row indicates a *CYP2D6* genotype. The mean and the median log MR values are colored based on intensity (cold to warm color). * No., number of subjects.

were *CYP2D6**1/*10*B*, *10*B*/*10*B*, and *1/*1. The genotypes *CYP2D6**1/*14 and *CYP2D6**1/*21 each occurred in one subject only and were unique to the validation set.

A dendrogram was used to visualize the relationship between genotypes and phenotypes (Fig. 1). *CYP2D6**5/*5, a known PM

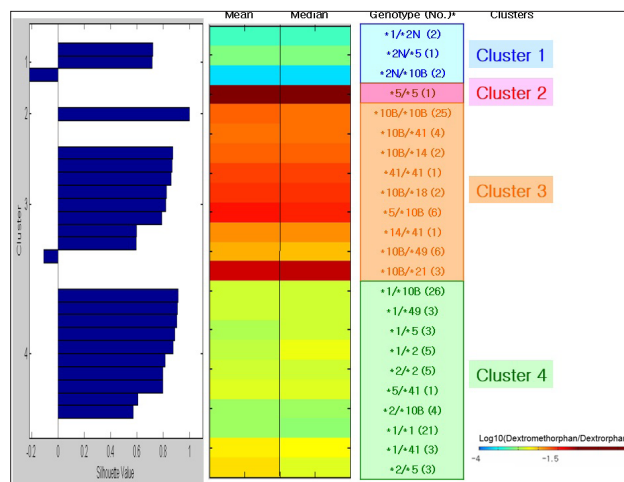


Figure 3. Color visualization and silhouette values of log-transformed urinary metabolic ratios (MR, dextromethorphan/dextrorphan) using the k-means cluster analysis based on four clusters (n=130). Each row indicates a *CYP2D6* genotype. The mean and the median log MR values are colored based on intensity (cold to warm color). * No., number of subjects.

type of *CYP2D6*, had the lowest observed enzyme activity and was represented by the color brown. The genotypes with the highest enzyme activity were *CYP2D6**2*N*/*10*B* and *1/*2*N*. The remaining genotypes were divided into two groups in the dendrogram, one group with higher levels of enzyme activity and the other with lower levels of activity. The higher activity group contained *1, *2, and *2*N* genotypes, which are known to have normal or increased enzyme activity. The lower activity group contained non-functional alleles (*5 or *21) or alleles known to have decreased activity (*10, *14, *18, and *41).

Based on the hierarchical clustering results, the number of the phenotype cluster was set between 2 and 10 for the k-means analysis. The number of the phenotype clusters with the highest mean silhouette value (average = 0.79) was three and the second highest (average = 0.71) was four. Color visualizations and silhouette values are presented in Figs. 2 and 3. The first genotype listed in each cluster was closest to the central value of the cluster, and the last genotype was furthest from the central value (and less characteristic of that cluster). Group numbers (Cluster 1, Cluster 2) were randomly assigned. *CYP2D6**5/*5, when classified into three clusters and four clusters by the k-means method, were classified as one separated cluster (Cluster 2) and *CYP2D6**1/*2*N*, *2*N*/*5, and *2*N*/*10*B* were in a cluster (Cluster 1) of four clusters. With the exception of *CYP2D6**5/*41, one cluster (Cluster 4) contained genotypes with at least one normal or increased function gene and the other cluster (Cluster 3) did not contain these genes in the four clusters. It was similar in the three clusters. The three clusters are shown in Fig. 2 and listed in Table 1. The information was easier and faster to interpret using color visualization. The mean log MR value for each genotype

Table 1. Results of the k-means cluster analysis based on three clusters of CYP2D6 genotype and phenotype data (n=130)

Genotype	Mean ^a	Median ^a	No ^b	Clusters
*2/*10B	-2.72	-2.72	4	1
*1/*1	-2.73	-2.75	21	1
*1/*5	-2.61	-2.48	3	1
*1/*49	-2.50	-2.51	3	1
*1/*10B	-2.47	-2.53	26	1
*1/*2	-2.54	-2.40	5	1
*2N/*5	-2.82	-2.82	1	1
*5/*41	-2.46	-2.46	1	1
*2/*2	-2.49	-2.48	5	1
*1/*41	-2.22	-2.28	3	1
*2/*5	-2.18	-2.45	3	1
*1/*2N	-3.08	-3.08	2	1
*2N/*10B	-3.47	-3.47	2	1
*5/*5	0.07	0.07	1	2
*10B/*10B	-1.62	-1.69	25	3
*10B/*41	-1.73	-1.72	4	3
*10B/*14	-1.65	-1.65	2	3
*41/*41	-1.53	-1.53	1	3
*10B/*18	-1.42	-1.42	2	3
*5/*10B	-1.31	-1.38	6	3
*14/*41	-1.83	-1.83	1	3
*10B/*49	-1.97	-2.00	6	3
*10B/*21	-1.03	-1.00	3	3

^amean and median of log-transformed urinary metabolic ratios (dextromethorphan/dextrorphan) for each genotype, ^bnumber of subjects.

ranged from -3.47 to 0.07, while the range for all subjects was -3.53 to 0.07.

To verify the results of the clustering analysis, phenotype groups in the validation set were determined using the ROC curve cutoff value of three and four clusters from the clustering set. The cluster with the *CYP2D6**5/*5 genotype, which included only one subject with the lowest level of enzyme activity, differed significantly from the other groups in terms of the cutoff value. There were no individuals with the *CYP2D6**5/*5 genotype among the 71 subjects in the validation set, and cutoff values were determined as one for three clusters and two for four clusters. The cutoff value in the three clusters was -2.053 with 90.0% sensitivity and 92.4% specificity. The cutoff values of the four clusters were -2.9668 and -2.053. Sensitivity and specificity values were 90.5% and 80.0%, respectively, for the former cutoff and 90.0% and 91.9%, respectively, for the latter cutoff. The concordance rate was 91.5% for one cutoff of the three clusters, and 85.3% for two cutoffs of the four clusters for the clustering

set. When the cutoff values were applied to the validation set, the concordance rate was 92.8% for the three clusters and 73.9% for the four clusters. A lower concordance rate was observed in both the clustering and validation data for the four clusters, and one cutoff value in the three clusters was equal to one in the four clusters. Genotypes included in two clusters (Cluster 3) were the same in the cluster with the lowest enzymatic activity, except the cluster containing *CYP2D6**5/*5 of the three and the four clusters. Concordance rates in both cases were high, but they were lower in the clusters with the highest (Cluster 1) and second highest (Cluster 4) enzyme activities in the four clusters.

Discussion

CYP2D6 phenotype classification based on genotype was determined by categorization into four groups (UM to PM according to functionality of alleles), by choosing an antimode using a frequency histogram or a probit plot of MR. Using an antimode to separate NM (EM) and IM groups was not feasible, and its use was limited to separating the PM group from other groups.[13-15]

With regard to function, PM was defined as a gene in which both alleles are nonfunctional (*5/*5), and IM (*5/*10) as having a nonfunctional and a decreased functional allele. UM and NM (EM) were classified as carrying at least three functional genes (*1/*2N) and having one or two functional genes (*1/*1), respectively.[16,17] To predict *CYP2D6* phenotype, a combination of genetic variations were selected as markers and four data mining techniques were employed to predict IM and PM groups in Hong Kong-Chinese data.[18] Two pieces of genetic information, *188C>T* and deletion genes, were sufficient to predict IM and PM groups that had clinical significance. The *188C>T* mutation was not previously identified in a study of 51 Korean subjects using direct DNA sequencing,[11] suggesting that different ethnic groups within Asian populations are genetically diverse. Previous study with the Hong Kong-Chinese data had applied the supervised method for *CYP2D6* phenotype classification based on known genetic function without phenotype data, which excludes the influence of other genes or environmental factors. Considering the effects of specific genotypes and environmental factors on drug metabolism, it is necessary to classify these genotypes based on phenotypes. Use of genetic information to predict phenotype, as was done in the present study, provides a more accurate assessment.

Gaedigk et al. divided groups into allele scores using the activity score (AS) system, which assigns scores according to the relative functionality of each allele.[19] The possible distribution of DM MR in each score, from 0 to over 2, was presented. Phenotype groups were further defined as slower EM (EM-S), faster EM (EM-F), PM, IM, EM, and UM using arbitrary antimode. Ethnic differences were identified as playing an important role, and the results could not be generalized to all ethnicities. Factors such as food and other genes may also affect *CYP2D6* enzyme activity. An advantage of the AS system is that the char-

acteristics of the genotype can be regarded as the distribution of enzyme activity; however, the disadvantage is that information on the function of each gene must precede it. Recent studies have emphasized the need to standardize CYP2D6 phenotype assignment from genotype, proposing the AS system as a more suitable option than the traditional classification method (four groups from UM to PM).[20] The AS system has been used to predict phenotypes from *CYP2D6* genotype across major populations using data from previous research.[21]

This study classified CYP2D6 phenotype for each genotype using an unsupervised method without considering the known genotype classification. The phenotype features extracted from each genotype were the mean, median, and SD of DM log MR values of subjects with the same genotype. More accurate results could be obtained with a larger number of subjects representing each genotype. A single subject may yield results that are inconsistent with other studies. It was possible to classify low frequency genotypes, including a single subject, in our study. The cluster with the lowest enzyme activity consisted of non-functional alleles and the cluster with the highest enzyme activity contained alleles with amplified function (i.e., *CYP2D6*2N*); similar findings have been observed in previous reports.[17, 22] Color visualization of the results allowed for rapid and easy identification of the characteristics of each genotype. Users are not required of knowledge on applied clustering algorithm, making the tool usable for clinicians and other researchers.

K-means clustering analysis was applied to the extracted phenotype features. The genotypes of the clustering and validation sets were different, and clustering results could not be applied directly to the validation data to evaluate accuracy. The cutoff value from the ROC curve, a log MR value, was derived from the two closest phenotype activity clusters obtained from the clustering set analysis. Concordance rates were calculated to determine whether genotypes and phenotypes from each individual in the validation set were consistent with the clusters established using the cutoff value. When classified into three clusters, the concordance rate was > 90% in both the clustering and validation sets, confirming that it could be used for clustering of other genotype-phenotype data. When divided into four clusters, the concordance rate was ~85% for the clustering set and ~74% for the validation set. It is possible that fewer subjects were sorted into the cluster with the highest level of activity of four clusters than in the group with three clusters, not effectively reflecting the features of each genotype. The three clusters were better in the aspect of silhouette values and concordance rates, but it would be more useful when divided into more than three clusters when considering usefulness for personalized therapy. Two subjects in the validation set had *CYP2D6*1/*14* and *CYP2D6*1/*21*, which were not observed in the clustering set. A more complete set that includes all genotypes may improve predictions of phenotype based on genotype. This type of cluster analysis can be used widely for ethnicities other than Koreans, if ethnicity-specific factors are considered.

CYP2D6 activity is influenced by ethnicity. High frequencies of *10 in Asian populations and *17 in African populations may influence phenotype patterns.[21,23] The frequency of *CYP2D6*10* in the Korean population has been reported as 45.6%.[11] When applying a DM log MR value of ≥ -0.5 (the PM cutoff proposed by Schmid et al.[15]), only *5/*5 was included in the data with a frequency of 0.8%. The findings of this study were consistent with the results of previous studies.

To predict CYP2D6 activity based on *CYP2D6* genotype, we analyzed genotype and phenotype data using two clustering algorithms and statistical analyses. The results were presented using color visualization, which allows for rapid identification of phenotype based on genotype. The clustering results and cutoff values were evaluated using untouched data, and relatively good concordance rates were observed. This technique can be applied to various drug metabolizing enzymes, such as other CYPs, for optimal personalized therapy.

Acknowledgements

This research was supported by a grant of Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI15C1537).

Conflicts of interests

-Authors: The authors have declared that no conflicts of interest exist regarding this study.

-Reviewers: Nothing to declare

-Editors: Nothing to declare

References

1. Ingelman-Sundberg M, Oscarson M, McLellan RA. Polymorphic human cytochrome P450 enzymes: An opportunity for individualized drug treatment. *Trends Pharmacol Sci* 1999;20:342-349.
2. Budnitz DS, Shehab N, Kegler SR, Richards CL. Medication use leading to emergency department visits for adverse drug events in older adults. *Ann Intern Med* 2007;147:755-765.
3. Ma MK, Woo MH, McLeod HL. Genetic basis of drug metabolism. *Am J Health Syst Pharm* 2002;59:2061-2069.
4. Nebert DW, Wikvall K, Miller WL. Human cytochromes P450 in health and disease. *Philos Trans R Soc Lond B Biol Sci* 2013;368:20120431. doi: 10.1098/rstb.2012.0431.
5. Cytochrome P450 homepage. <http://drnelson.uthsc.edu/cytochromeP450.html> Accessed 18 July 2017.
6. Ingelman-Sundberg M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): Clinical consequences, evolutionary aspects and functional diversity. *Pharmacogenomics J* 2005;5:6-13.
7. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* 2013;138:103-141. doi: 10.1016/j.pharmthera.2012.12.007.
8. Sim SC, Ingelman-Sundberg M. Update on allele nomenclature for human cytochromes P450 and the human cytochrome P450 allele (CYP-allele) nomenclature database. *Methods Mol Biol* 2013;987:251-259. doi: 10.1129/978-1-62703-321-3_21.
9. Bernard S, Neville KA, Nguyen AT, Flockhart DA. Interethnic differences in genetic polymorphisms of CYP2D6 in the U.S. Population: Clinical implications. *Oncologist* 2006;11:126-135.
10. Caudle KE, Dunnenberger HM, Freimuth RR, Peterson JF, Burlison JD,

- Whirl-Carrillo M, et al. Standardizing terms for clinical pharmacogenetic test results: Consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC). *Genet Med* 2017;19:215-223. doi: 10.1038/gim.2016.87.
11. Lee SJ, Lee SS, Jung HJ, Kim HS, Park SJ, Yeo CW, et al. Discovery of novel functional variants and extensive evaluation of CYP2D6 genetic polymorphisms in Koreans. *Drug Metab Dispos* 2009;37:1464-1470. doi: 10.1124/dmd.108.022368.
 12. Kim EY, Lee SS, Jung HJ, Jung HE, Yeo CW, Shon JH, et al. Robust CYP2D6 genotype assay including copy number variation using multiplex single-base extension for Asian populations. *Clin Chim Acta* 2010;411:2043-2048. doi: 10.1016/j.cca.2010.08.042.
 13. Gaedigk A, Gotschall RR, Forbes NS, Simon SD, Kearns GL, Leeder JS. Optimization of cytochrome P450 2D6 (CYP2D6) phenotype assignment using a genotyping algorithm based on allele frequency data. *Pharmacogenetics* 1999;9:669-682.
 14. Eichelbaum M, Woolhouse NM. Inter-ethnic difference in sparteine oxidation among Ghanaians and Germans. *Eur J Clin Pharmacol* 1985;28:79-83.
 15. Schmid B, Bircher J, Preisig R, Kupfer A. Polymorphic dextromethorphan metabolism: Co-segregation of oxidative O-demethylation with debrisoquin hydroxylation. *Clin Pharmacol Ther* 1985;38:618-624.
 16. Sachse C, Brockmüller J, Hildebrand M, Müller K, Roots L. Correctness of prediction of the CYP2D6 phenotype confirmed by genotyping 47 intermediate and poor metabolizers of debrisoquine. *Pharmacogenetics* 1998;8:181-185.
 17. Chou WH, Yan FX, Robbins-Weilert DK, Ryder TB, Liu WW, Perbost C, et al. Comparison of two CYP2D6 genotyping methods and assessment of genotype-phenotype relationships. *Clin Chem* 2003;49:542-551.
 18. Sabbagh A, Darlu P. Data-mining methods as useful tools for predicting individual drug response: Application to CYP2D6 data. *Hum Hered* 2006;62:119-134.
 19. Gaedigk A, Simon SD, Pearce RE, Bradford LD, Kennedy MJ, Leeder JS. The CYP2D6 activity score: Translating genotype information into a qualitative measure of phenotype. *Clin Pharmacol Ther* 2008;83:234-242.
 20. Hicks JK, Swen JJ, Gaedigk A. Challenges in CYP2D6 phenotype assignment from genotype data: A critical assessment and call for standardization. *Curr Drug Metab* 2014;15:218-232.
 21. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. *Genet Med* 2017;19:69-76. doi: 10.1038/gim.2016.80.
 22. Streetman DS, Ellis RE, Nafziger AN, Leeder JS, Gaedigk A, Gotschall R, et al. Dose dependency of dextromethorphan for cytochrome P450 2D6 (CYP2D6) phenotyping. *Clin Pharmacol Ther* 1999;66:535-541.
 23. Droll K, Bruce-Mensah K, Otton SV, Gaedigk A, Sellers EM, Tyndale RF. Comparison of three CYP2D6 probe substrates and genotype in Ghanaians, Chinese and Caucasians. *Pharmacogenetics* 1998;8:325-333.