



OPEN ACCESS

EDITED BY

Balram Chowbay,
National Cancer Centre Singapore,
Singapore

REVIEWED BY

Erika Cecchin,
Aviano Oncology Reference Center
(IRCCS), Italy
Nancy Hakooz,
The University of Jordan, Jordan

*CORRESPONDENCE

Michel S. Naslavsky,
✉ mnaslavsky@usp.br

RECEIVED 03 March 2023

ACCEPTED 10 April 2023

PUBLISHED 10 May 2023

CITATION

Bertholim-Nasciben L, Scliar MO,
Debortoli G, Thiruvahindrapuram B,
Scherer SW, Duarte YAO, Zatz M,
Suarez-Kurtz G, Parra EJ and
Naslavsky MS (2023), Characterization of
pharmacogenomic variants in a Brazilian
admixed cohort of elderly individuals
based on whole-genome
sequencing data.

Front. Pharmacol. 14:1178715.

doi: 10.3389/fphar.2023.1178715

COPYRIGHT

© 2023 Bertholim-Nasciben, Scliar,
Debortoli, Thiruvahindrapuram, Scherer,
Duarte, Zatz, Suarez-Kurtz, Parra and
Naslavsky. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Characterization of pharmacogenomic variants in a Brazilian admixed cohort of elderly individuals based on whole-genome sequencing data

Luciana Bertholim-Nasciben^{1,2}, Marilia O. Scliar²,
Guilherme Debortoli³, Bhooma Thiruvahindrapuram⁴,
Stephen W. Scherer^{4,5}, Yeda A. O. Duarte⁶, Mayana Zatz^{2,7},
Guilherme Suarez-Kurtz⁸, Esteban J. Parra³ and
Michel S. Naslavsky^{2,7,9*}

¹School of Public Health, University of São Paulo, São Paulo, Brazil, ²Human Genome and Stem Cell Research Center, University of São Paulo, São Paulo, Brazil, ³Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON, Canada, ⁴The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada, ⁵Department of Molecular Genetics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada, ⁶Medical-Surgical Nursing Department, School of Nursing, University of São Paulo, São Paulo, Brazil, ⁷Department of Genetics and Evolutionary Biology, Biosciences Institute, University of São Paulo, São Paulo, Brazil, ⁸Divisão de Pesquisa Clínica e Desenvolvimento Tecnológico, Instituto Nacional de Câncer, Rio de Janeiro, Brazil, ⁹Hospital Israelita Albert Einstein, São Paulo, Brazil

Introduction: Research in the field of pharmacogenomics (PGx) aims to identify genetic variants that modulate response to drugs, through alterations in their pharmacokinetics (PK) or pharmacodynamics (PD). The distribution of PGx variants differs considerably among populations, and whole-genome sequencing (WGS) plays a major role as a comprehensive approach to detect both common and rare variants. This study evaluated the frequency of PGx markers in the context of the Brazilian population, using data from a population-based admixed cohort from Sao Paulo, Brazil, which includes variants from WGS of 1,171 unrelated, elderly individuals.

Methods: The Stargazer tool was used to call star alleles and structural variants (SVs) from 38 pharmacogenes. Clinically relevant variants were investigated, and the predicted drug response phenotype was analyzed in combination with the medication record to assess individuals potentially at high-risk of gene-drug interaction.

Results: In total, 352 unique star alleles or haplotypes were observed, of which 255 and 199 had a frequency < 0.05 and < 0.01, respectively. For star alleles with frequency > 5% ($n = 97$), decreased, loss-of-function and unknown function accounted for 13.4%, 8.2% and 27.8% of alleles or haplotypes, respectively. Structural variants (SVs) were identified in 35 genes for at least one individual, and occurred with frequencies >5% for CYP2D6, CYP2A6, GSTM1, and UGT2B17. Overall 98.0% of the individuals carried at least one high risk genotype-predicted phenotype in pharmacogenes with PharmGKB level of evidence 1A for drug interaction. The Electronic Health Record (EHR) Priority Result Notation and the cohort medication registry were combined to assess high-risk gene-drug interactions. In general, 42.0% of the cohort used at least one PharmGKB evidence level 1A drug, and 18.9% of individuals who used PharmGKB evidence level 1A drugs had a genotype-predicted phenotype of high-risk gene-drug interaction.

Conclusion: This study described the applicability of next-generation sequencing (NGS) techniques for translating PGx variants into clinically relevant phenotypes on a large scale in the Brazilian population and explores the feasibility of systematic adoption of PGx testing in Brazil.

KEYWORDS

pharmacogenomics, admixture, population cohort, whole-genome sequencing, PharmGKB, CPIC guidelines

Introduction

The concept of precision medicine is based on identifying individuals at risk of developing diseases, their trajectories, and their individual response to treatments. Pharmacogenomics (PGx) plays an important role in precision medicine and deals with interindividual variation of drug response due to genetic variants across the genome, which affect the efficacy and toxicity of drugs through alterations in their pharmacokinetics (PK)—absorption, bioavailability, distribution, metabolism, and excretion—or pharmacodynamics (PD) (Relling and Evans, 2015).

The Pharmacogenomics Knowledgebase (PharmGKB¹) is a major partner in PGx research and implementation in clinical practice, through the collection of primary PGx data, curation, and annotation of peer-reviewed literature on gene–drug associations (Barbarino et al., 2018; Whirl-Carrillo et al., 2021). Presently, there are more than 800 drugs in the PharmGKB database, for which associations with genetic variants have been reported². However, only a limited fraction of these associations, comprising 148 drugs, has been translated into genotype-based dosing recommendations by PGx-focused independent initiatives such as the Clinical Pharmacogenetics Implementation Consortium (CPIC) (Relling et al., 2020).

Several factors play a role in the translation of PGx findings into the clinic and objective criteria such as “levels of evidence” from PharmGKB which are considered by the CPIC as the main standard to classify these findings. The level of evidence of each gene–drug relationship is determined by the parameters of the study, such as strength of association, effect size, cohort size, and the reproducibility of the results³. It is important to note that genomic diversity and variation play a major role, since the distribution of PGx variants (both occurrence of rare variants and frequency of common polymorphisms) and the extent of linkage disequilibrium (LD) differ considerably among populations, with important implications for design of clinical trials and genome-wide association studies (GWASs) (Suarez-Kurtz and Parra, 2018).

For Brazilians and other Latin American (LA) populations, distinct patterns of admixture between different continental groups create additional challenges to PGx implementation in clinical practice (Suarez-Kurtz and Pena, 2007; Suarez-Kurtz and Parra, 2018). Accordingly, standardized guidelines to inform and adjust prescriptions based on PGx data require deeper knowledge of the

frequency and effect size of PGx variants across Latin American populations. In this context, whole-genome sequencing (WGS) plays a major role as a comprehensive approach to detect both common and rare variants, and in the development of algorithms to predict the functionality of rare variants (Suarez-Kurtz and Parra, 2018).

This study assessed PGx markers in the Brazilian population, using data from the “Health, Well-Being, and Aging Study” (SABE—*Saúde, Bem-estar e Envelhecimento*), a population-based cohort from Sao Paulo, Brazil, which includes variants from whole-genome sequences of 1,171 unrelated individuals. The frequency of star alleles of pharmacogenes was investigated, and the predicted phenotypes were analyzed in combination with medication records for the assessment of individuals potentially at high-risk for gene–drug interactions.

Materials and methods

SABE project and the study population

The SABE study cohorts comprise population-based probability samples of individuals aged 60 years and older and were designed to provide information on health indicators of the elderly population in the city of São Paulo, Brazil, through comprehensive in-home interviews and biological sample collection every 5 years. This cross-sectional study included individuals who participated in the third wave of data collection in 2010. A detailed description of the project under the coordination of the School of Public Health at the University of São Paulo (FSP-USP) can be found elsewhere (Lebrão and Laurenti, 2005; Naslavsky et al., 2022). Individuals responded to a long questionnaire which included questions related to self-reported health conditions, such as hypertension, heart and cardiovascular conditions, diabetes, cancer, chronic pulmonary disorders, joint conditions, osteoporosis, anemia, and depression. One section of the questionnaire is devoted to collect information about medication and supplement intake. Biological samples were collected, including peripheral blood for DNA biobanking.

The SABE project was approved by the FSP-USP Institutional Review Board and the National Committee of Ethics in Research. All participants signed consent forms according to the Brazilian regulatory requirements human research.

Variant discovery

A total of 1,335 SABE participants were enrolled in the 2010 round of data collection. DNA was extracted from these individuals, and 1,200 DNA samples met the quality criteria and

1 www.pharmgkb.org

2 www.pharmgkb.org/downloads

3 www.pharmgkb.org/page/varAnnScoring

were submitted to WGS. Further details on the adopted methodology can be found in the WGS flagship publication (Naslavsky et al., 2022). Briefly, Illumina HiSeqX sequencers were used with a 30X target coverage and a 150 base paired-end single index read format. Relatedness was assessed by the KING toolset (Manichaikul et al., 2010), and to avoid inflating population frequencies for rare alleles, only one individual (proband) was maintained when identifying siblings, duos, or other pairs of up to three degrees of relationships. GATK flags (Auwera et al., 2013) and an in-house genotype and variant flagging algorithm were applied to filter out low-quality variants, as described in Naslavsky et al. (2022). The variants and allelic frequencies of the aggregated sample of 1,171 unrelated individuals submitted to WGS are publicly available on the ABraOM (*Arquivo Brasileiro Online de Mutações*) platform⁴. The frequency of clinically important variants in *CACNASIS*, *CFTR*, *DPYD*, *IFNL3*, and *RYR1* were extracted from the ABraOM dataset since these genes are not represented in the Stargazer pipeline.

Pharmacogenes star allele assessment

WGS data had been previously mapped to human reference GRCh38 using ISIS analysis software (Raczy et al., 2013), and reads overlapping with the 38 pharmacogenes of interest regions were extracted from BAM files, converted to FastQ files, and realigned to hg19 using BWA-MEM v0.7.12 (Li and Durbin, 2009). Duplicate reads were removed using MarkDuplicates from Picard v1.79. GATK3.7 (Auwera et al., 2013) was used for indel realignment, base quality score recalibration, and variant joint calling. The following hard filters were applied to exclude low-quality SNPs (QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, and SOR > 3.0) and indels (QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0, and SOR > 10.0). Read-depth files were obtained through the GATK's DepthOfCoverage function with mapping and base quality thresholds of 20 or greater.

The star alleles of 38 pharmacogenes were called using the tool “genotype” from Stargazer v1.0.7 genotyping pipeline (Lee et al., 2019a; Lee et al., 2019b), which uses the hard-filtered vcf and coverage files as input, the program Beagle (Browning et al., 2018), and the 1000 Genomes Project (Auton et al., 2015) haplotype as a reference panel for phasing. Phased SNVs and indels were then matched to star alleles. Stargazer used the read depth from the coverage file to convert to copy number by performing intrasample normalization using read depth from a control GDF file (Lee et al., 2019a). Diplotypes are defined as two alleles or haplotypes carried by a given individual.

Predicted phenotype assignment

For pharmacogenes with PharmGKB evidence level 1A, PharmGKB's diplotype-phenotype translation tables⁵ were used

to map each individual diplotype to a predicted phenotype and calculate phenotype frequencies (Supplementary Figure S1). Diplotypes that were not listed in translation tables were assigned as unknown functions, with the exception of *CYP2D6*, for which several diplotypes present in our sample were missing in PharmGKB's translation tables, and the sum of the activity score (AS) for each allele was used to assign the predicted metabolic phenotype according to the CPIC guidelines (Caudle et al., 2020).

Furthermore, predicted phenotypes were matched to the Electronic Health Record (EHR) Priority Result Notation (risk)⁵, and the frequency of individuals at high-risk of gene-drug interaction was verified.

Since diplotype-phenotype translation tables have not been reported for *CYP4F2* and *VKORC1*, the assignment of predicted phenotypes was interpreted from the Warfarin dosing guideline (Johnson et al., 2017). *CYP4F2**3 (c.1297G>A; p. Val433Met; rs2108622) is listed as a decreased function allele and individuals with one or two copies of *CYP4F2**3 were assigned as higher warfarin dose phenotype. *VKORC1* (c.-1639G>A, rs9923231) is associated with warfarin sensitivity, and patients with one or two -1639A require progressively lower warfarin doses. Thus, individuals with one or two copies of *VKORC1**2 (rs9923231, and the linked rs9934438 and rs235961) were assigned as decreased warfarin dose phenotype.

Medication analysis

The SABE in-home interview includes the collection of information about the medication that respondents were taking at the time of the interview. The medication section from the 2010 round of data collection was first checked against a list of 45 PharmGKB evidence level 1A drug-gene pairs (Supplementary Figure S1) to obtain the fraction of the cohort who has taken one or more of the drugs listed. Then, potential PGx interactions were verified by checking the individuals that had any predicted phenotype designated as high-risk following EHR Priority Result Notation and had also taken any specific drug associated with the high-risk phenotype.

Results

Samples

A total of 1,171 unrelated, elderly Brazilian individuals from the SABE 2010 cohort passed sample-level QC criteria, including 427 (36.5%) men and 744 (63.5%) women, with a median age of 71 years (IQR = 64–80). The distribution of self-reported color/race categories, according to the Brazilian census was “White,” 58.1% ($n = 680$); “Brown” (*Pardo* in Brazilian Portuguese), 28.2% ($n = 330$); “Black,” 6.4% ($n = 75$); “Yellow” (referring to Asian extraction), 2.7% ($n = 32$); others, 2.1% ($n = 25$); and no answer, 2.5% ($n = 29$). It is important to note that these census “color/race” classifications are sociopolitical and not biological constructs. In Brazil, multiple studies have highlighted the complex relationships between “color/race” and genetic ancestry (e.g., estimates of the geographical origin of the ancestors of the individuals in a sample

⁴ abraom.ib.usp.br

⁵ www.pharmgkb.org/page/pgxGeneRef

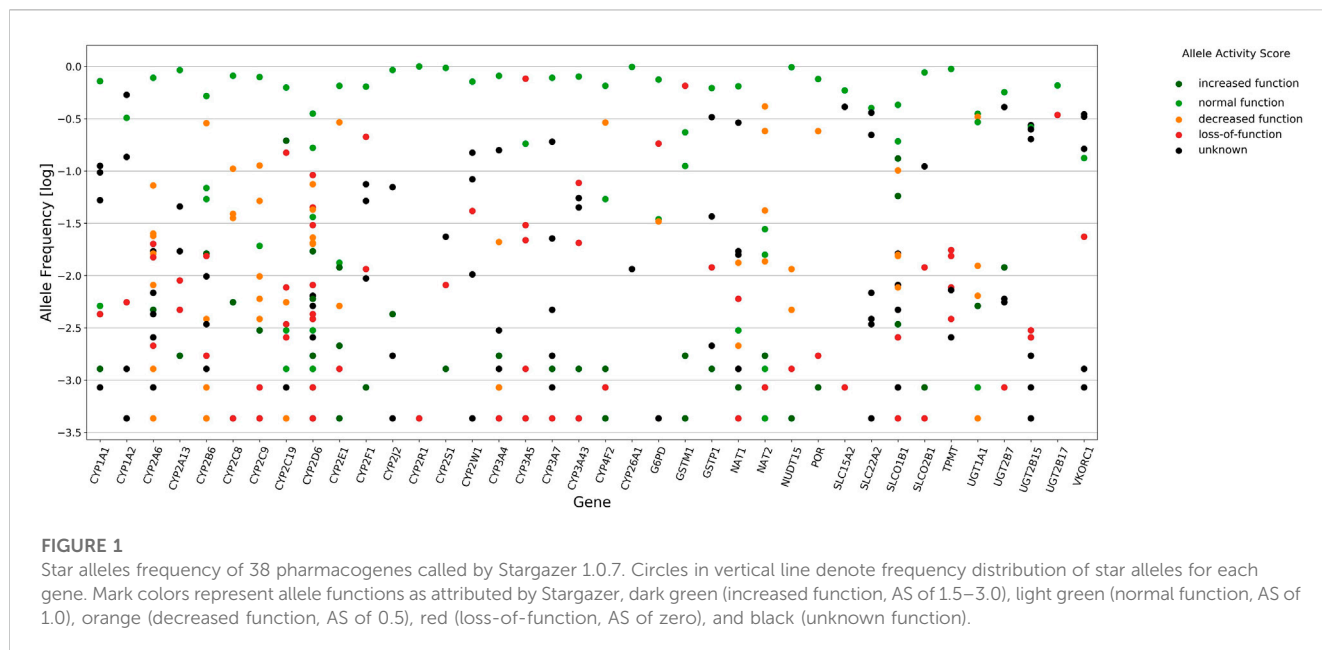


FIGURE 1

Star alleles frequency of 38 pharmacogenes called by Stargazer 1.0.7. Circles in vertical line denote frequency distribution of star alleles for each gene. Mark colors represent allele functions as attributed by Stargazer, dark green (increased function, AS of 1.5–3.0), light green (normal function, AS of 1.0), orange (decreased function, AS of 0.5), red (loss-of-function, AS of zero), and black (unknown function).

based on statistical inferences made from genomic data) (Pena et al., 2011; Naslavsky et al., 2022), and we provide a relevant example when discussing the distribution of *CYP2C9* variants in one of the sections of this article. We would like to note that when writing self-reported “White or Black” persons, we are referring to the Brazilian census categories.

Star alleles assessment

Stargazer 1.0.7 was used to call star alleles in 38 pharmacogenes in the study cohort, and the frequency of haplotypes and diplotypes are presented in [Supplementary Tables S1, 2](#), respectively. In total, 352 unique star alleles or haplotypes were observed in all pharmacogenes assessed. Among these, 255 and 199 had a frequency <5% and <1%, respectively. For star alleles with frequency >5% ($n = 97$), decreased function, loss-of-function, and unknown function accounted for 13.4%, 8.2%, and 27.8% of alleles or haplotypes, respectively (Figure 1). Structural variants (SVs) were identified in 35 genes for at least one individual, although the allele frequencies of SVs were >5% only for *CYP2D6*, *CYP2A6*, *GSTM1*, and *UGT2B17*. A total of 103 different duplications/deletions of star alleles and 16 different rearrangements/deletions in *CYP2A6*, *CYP2B6*, *CYP2D6*, *CYP2E1*, *SLC22A2*, and *UGT2B15* were identified. The allele frequency of complete deletion of *GSTM1* and *UGT2B17* was 65.2% and 34.3%, respectively (Supplementary Table S1).

CYP2D6 showed the highest levels of polymorphism among all genes, with 43 identified unique star alleles (Figure 1), and 23.9% of individuals carrying SVs including rearrangements with loss-of-function (*68+*4, *4N+*4, *13), decreased function (*36+*10) and normal function (*S1+*1 and *83+*2), and copy number variation (CNV) ranging from zero to three gene copies. Among alleles that have not been reported in

PharmVar, we detected *CYP2D6* *21 × 2 and *83+*2 and two new star alleles referred to as *S1+*1 and *S2+*1 (Supplementary Figure S2).

A total of 21 *CYP* genes were analyzed, and the highest frequency of loss-of-function and decreased function alleles, considering the AS of 0 or 0.5 designated by Stargazer, occurred in *CYP3A5* (81.7%), followed by *CYP2D6* (38.6%), *CYP2B6* (30.9%), *CYP2E1* (29.8%), *CYP4F2* (29.2%), *CYP2F1* (22.3%), *CYP2C9* (18.6%), *CYP2A6*, and *CYP2C8* (18.5%). The frequency of loss-of-function and decreased function alleles was less than 10% in *CYP2W1*, *CYP3A4*, *CYP2A13*, *CYP2S1*, *CYP1A2*, *CYP1A1*, *CYP2J2*, *CYP3A7*, *CYP2R1*, and *CYP3A43*.

The frequency of *NAT2* decreased function star alleles was 71.0%. In the case of *NAT1*, normal function and unknown function alleles accounted for 64.6% and 32.5%, respectively. *POR**28 (decreased function) had a frequency of 24.0%. The combined frequency of decreased function *NUDT15**3 and *NUDT15**4 was 1.6%.

Four solute carrier (SLC) transporter genes were studied (*SLCO1B1*, *SLCO2B1*, *SLC15A2*, and *SLC22A2*), and *SLCO1B1* showed the highest degree of polymorphism with 20 identified star alleles, 15.5% alleles with loss-of-function or decreased function, and 19.3% of alleles with increased function.

The combined frequency of *TPMT* loss-of-function alleles (*2, *3A, *3B, and *3C) was 4.4%. Four UDP-glucuronosyltransferase (UGT) phase II metabolism enzymes were evaluated. *UGT1A1**28 is the most common decreased function allele and had a frequency of 32.9%. The decreased function allele *UGT1A1**37 had a frequency of 1.2%, while *UGT1A1**6 and *UGT1A1**7 were detected at a frequency <0.1%.

Most of the star alleles from *VKORC1* were designated by Stargazer as “unknown function” (84.4%), including *VKORC1**2 defined by the variant rs9923231 (and the linked rs9934438 and rs235961), related to warfarin dosage, which has a frequency of 33.1%.

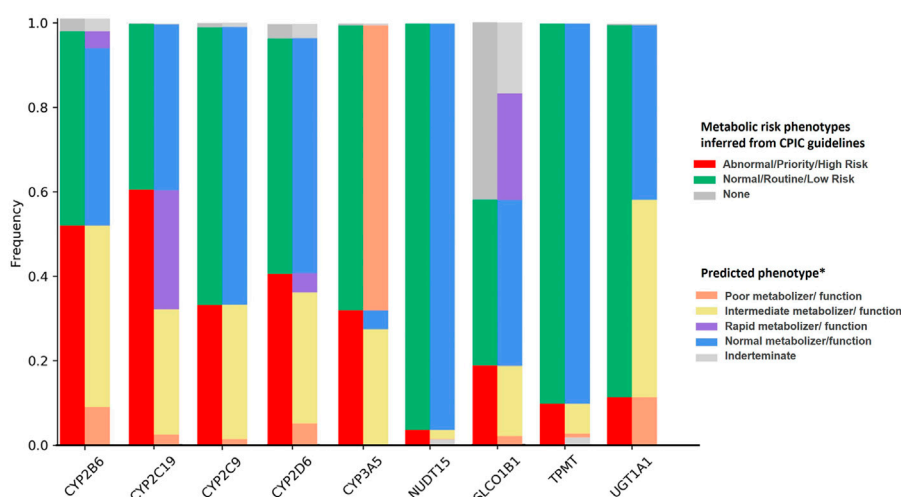


FIGURE 2

Frequency of predicted metabolizer/function phenotypes and gene–drug priority risk categories based on the CPIC guidelines for pharmacogenes with PharmGKB level of evidence 1A for drug interaction. Bars on the left represent phenotype risk using Electronic Health Record Priority Result Notation. Bars on the right indicate the predicted phenotype obtained by matching diplotypes assigned by Stargazer to phenotypes using PharmGKB/CPIC diplotype–phenotype reference tables [Poor/Intermediate/Normal/Rapid] function refers to SLCO1B1 phenotypes. *Predicted phenotype categories were grouped as follows: Intermediate metabolizer/function also includes: CYP2C19 Likely Intermediate Metabolizer (CYP3A5/NUDT15/TPMT) Possible Intermediate Metabolizer, SLCO1B1 Decreased function, and SLCO1B1 Possible Decreased Function; poor metabolizer/function also includes: CYP2C19 Likely Poor Metabolizer, and SLCO1B1 Possible Poor Function; and rapid metabolizer/function also includes SLCO1B1 Possible Increased Function, and CYP2D6 Ultrarapid Metabolizer.

Variant analysis of *CACNAS1S*, *CFTR*, *DPYD*, *IFNL3*, and *RYR1*

In addition to the evaluation of star alleles, the SABE cohort WGS dataset (deposited in ABraOM) was used to verify the frequency of clinically important variants in *CACNAS1S*, *CFTR*, *DPYD*, *IFNL3*, and *RYR1* (Supplementary Table S3). No *CACNAS1S* or *RYR1* actionable variants were found, and three individuals were identified with the rare loss-of-function *DPYD* variant rs3918290 in heterozygosis.

Clinically actionable pharmacogenes

To estimate the possible clinical impact of the findings, the predicted phenotype was obtained by analyzing the genotype data of 1,171 individuals for pharmacogenes with PharmGKB evidence level 1A. For this, PharmGKB reference translation tables⁵ were used to map each individual diplotype to a predicted phenotype and then to individual EHR priority result notation (Supplementary Figure S1), which is used to predict individuals at potential risk of an adverse or untoward response to medications due to the gene–drug interaction (Figure 2). Overall, 98.0% of the individuals carried at least one high-risk genotype in nine genes analyzed.

A total of 49 diplotypes were identified in *CYP2B6*, half of which were associated with intermediate or poor metabolizer phenotypes and classified as high risk for an adverse or poor response to medications that are metabolized by *CYP2B6* (51.6% of the individuals). The frequency of diplotypes with one or two copies of *CYP2B6**6, associated with lower protein expression and activity, was 45.8%.

In the case of *CYP2C19*, carriers of one or two copies of *17, *3, and *2 are considered at potential risk of an adverse or poor response to medications metabolized by the gene. Due to the high frequency of those diplotypes in the cohort, mainly *CYP2C19* *1/*17 (24.3%), *1/*2 (19.6%), and *2/*17 (6.4%), *CYP2C19* had the highest frequency of genotypes associated with actionable phenotypes.

According to the CPIC, individuals with intermediate and poor metabolizer *CYP2C9* phenotypes are considered to be at high risk for adverse reactions to medications that are affected by *CYP2C9*. The intermediate metabolizer diplotypes *CYP2C9* *1/*2 and *1/*3 were present in 17.5% and 8.4% of individuals, respectively, and a total of 33.2% were classified into the high-risk EHR annotation group.

The frequency of individuals at high risk for an adverse or untoward response to medication metabolized by *CYP2D6* was 40.8%, including ultrarapid (4.6%), intermediate (31.3%), and poor metabolizer phenotypes (5.0%). The most frequent high-risk diplotypes were *CYP2D6* *1/*4 (6.15%), *1/*68*4 (3.59%), *1/*5 (2.56%), and *2/*4 (2.56%), all intermediate metabolizers.

CYP3A5 high-risk phenotypes, which are normal and intermediate metabolizers, were verified in 4.5% and 27.4% of our cohort, respectively, and consist of diplotypes of one or two copies of *CYP3A5* *1.

Less than 20% of the individuals were found to be at high risk of gene–drug interactions due to variants in *NUDT15*, *SLCO1B1*, *TPMT*, and *UGT1A1*. For *NUDT15*, intermediate metabolizer (*1/*3) is considered high-risk phenotypes and occurred in 2.1% of individuals. The frequency of individuals with *SLCO1B1* high-risk phenotypes (decreased function or poor function) was 18.7%, including individuals with one or two copies of

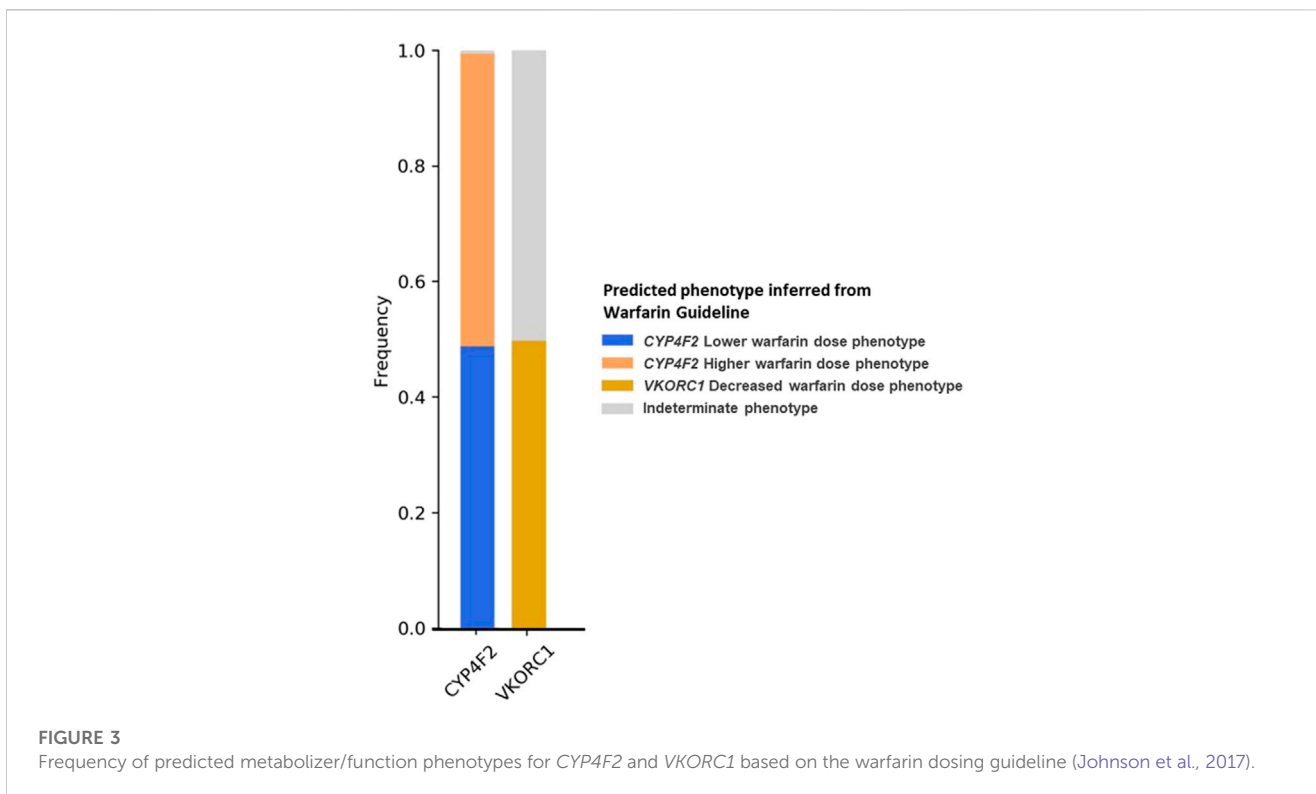


TABLE 1 Cohort medication use and predicted high-risk individuals by pharmacogene.

Gene	Drugs ^a	% Taking drugs (n)	% Taking drugs at high risk (n) ^b
<i>CYP2C19</i>	Amitriptyline, citalopram, clomipramine, clopidogrel, escitalopram, imipramine, lansoprazole, omeprazole, pantoprazole, and sertraline	25.1 (278)	14.5 (161)
<i>SLCO1B1</i>	Simvastatin	20.1 (223)	3.0 (33)
<i>CYP2D6</i>	Amitriptyline, clomipramine, codeine, fluvoxamine, imipramine, metoprolol, nortriptyline, paroxetine, propafenone, risperidone, tramadol, and venlafaxine	6.7 (74)	2.4 (27)
<i>CYP2C9</i>	Celecoxib, ibuprofen, meloxicam, phenytoin, piroxicam, tenoxicam, and warfarin	4.7 (52)	1.5 (17)
<i>CYP4F2/VKORC1</i>	Warfarin	1.4 (15)	0.5 (5)
<i>CYP2B6</i>	Efavirenz	0.2 (2)	0.1 (1)

^aDrugs identified in the cohort among 45 PharmGKB evidence level 1A drug–gene pairs (Supplementary Figure S1).

^bIndividuals whose predicted phenotype was designated as high-risk following the EHR Priority Result Notation and has also taken any specific drug associated with the high-risk phenotype.

SLCO1B1 *5, *15, or *17. *TPMT* intermediate metabolizer, poor metabolizer, and indeterminate metabolizer phenotypes are considered to be at high risk of gene–drug interactions and consisted of 9.8% of individuals. The frequency of individuals with *UGT1A1* *28/*28 (poor metabolizer predicted phenotype) was 11.3%, which is the only high-risk *UGT1A1* diplotype identified in our cohort.

CYP4F2 and *VKORC1* had the EHR Priority Result Notation designated according to warfarin guidelines. Therefore, the *CYP4F2* lower warfarin dose phenotype individuals (carriers of one or two copies of *CYP4F2**3) were 48.8% of our cohort, and *VKORC1* decreased

warfarin dose phenotype (carriers of one or two copies of *VKORC1**2), which accounted for 49.8% of our cohort (Figure 3).

Real-world medication use

The analysis of medication use was performed to predict the clinical impact of the findings in a real-world scenario. For this, the PharmGKB evidence level 1A medication used in outpatient settings was verified in the SABE questionnaire. The median number of drugs used was four (IQR = 2–6). In general, 42.0% of the individuals

used at least one PharmGKB level 1A drug, 18.9% were at potentially high risk for adverse effects, and 2.7% had more than one PGx interaction. Simvastatin and omeprazole were in use by 20.1 and 19.5% of the cohort, amitriptyline by 3.3%, sertraline by 2.3%, and ibuprofen by 2.2%. All other drugs were taken by <1.5% of the cohort (Table 1 and Supplementary Table S4).

Discussion

Genetic variation plays an important role on drug response, and the scientific community has been making efforts in the last few decades to translate PGx advances into the clinic. A better understanding of human genomic diversity is necessary to broaden the scope of clinical guidelines and implementation of PGx-informed prescription and also has important implications for the design of clinical trials. The present study uncovers the frequency of star alleles and predicted phenotypes of 38 pharmacogenes from WGS data for a population-based sample containing more than 1,000 individuals of admixed ancestry from Brazil (Naslavsky et al., 2022), with a distinctive demographic history as compared to other Latin American populations (Popejoy and Fullerton, 2016; Sirugo et al., 2019). Available medication data were used to assess potential pharmacogenes–drug interactions.

In general, the frequencies of the star alleles identified in the SABE WGS cohort were in accordance with previous studies in the Brazilian population (Rodrigues-Soares et al., 2018; Rodrigues-Soares et al., 2020), although a higher number of star alleles with a frequency <5% ($n = 255$) and <1% ($n = 199$) have been identified in our study. Genotyping technologies are able to identify the most frequent variants but lack the ability to reveal rare and deleterious ones that have been proven to play a significant role in the field of PGx (Ingelman-Sundberg et al., 2018; De Mattia et al., 2022; Gray et al., 2022). An analysis of the UK Biobank sample, which included nearly 50,000 subjects with both imputed data from genotyping arrays and exome sequencing, compared their ability to call haplotypes and phenotypes in 14 clinically important pharmacogenes. Despite the high concordance between techniques for most genes, the analysis revealed extremely low concordance for highly polymorphic pharmacogenes such as *CYP2D6*, where imputed data from genotyping arrays may not capture the wide range of variation (McInnes et al., 2021).

Several initiatives have used NGS to analyze PGx markers in diverse populations, using large genomic databases (Reisberg et al., 2019; Luo et al., 2021; McInnes et al., 2021; Taliun et al., 2021) or databases from specific populations (Cohn et al., 2017; Al-Mahayri et al., 2020; Mauleekoonphairoj et al., 2020). The advances in genomic analyses by NGS have enabled scientists to investigate the contribution of rare variants to complex diseases and PGx. Recently, Ingelman-Sundberg et al. (2018) analyzed the distribution of rare and common variants in 208 pharmacogenes by analyzing the exome sequencing data from 60,706 unrelated individuals from ExAC. They reported that the vast majority of variants were rare (98.5%; MAF <1%) or very rare (96.2%; MAF <0.1%), and for those variants, there was a strong enrichment in consequences predicted to cause functional alterations, suggesting that a substantial part of

the unexplained interindividual differences in drug metabolism phenotypes can be attributed to rare genetic variants.

NGS-based techniques have also facilitated the large-scale detection of SVs (i.e., microscopic or submicroscopic genomic alterations comprising DNA segments larger than 50 bp), including CNVs (duplications or deletions of DNA segments), translocations, inversions, and combinations of the same (Feuk et al., 2006; Trost et al., 2018). Numerous algorithms have been developed for detecting SVs from large-scale sequencing data, and they differ in both sensitivity and specificity mainly due to differences in SV-related and sequencing library-related properties, leading to variable depth of coverage (Guan and Sung, 2016). This heterogeneity motivated the development of best practices for the detection of germline CNVs, which represent the majority and the most clinically significant type of SVs, from short-read WGS data (Trost et al., 2018), less prone to variability in sequencing depth of coverage.

SVs have been shown to play a clear role in the field of PGx (Johansson and Ingelman-Sundberg, 2009; He et al., 2011; Santos et al., 2018) and were originally described in *CYP2D6*, which metabolizes around 25% of all drugs in clinical use, and has been extensively studied to uncover the functionality of its multiple variants and haplotypes including CNVs and other complex SVs that are not detected by conventional techniques. Its homology to the pseudogenes, *CYP2D7* and *CYP2D8*, becomes a challenge for the interrogation by short-read NGS (Schwarz et al., 2019; Luo et al., 2021). Recently, specific bioinformatics algorithms have been developed for pharmacogenes genotyping based on high-throughput sequencing data, most of them using *CYP2D6* as a model, such as Stargazer (Lee et al., 2019a; Lee et al., 2019b), Astrolabe (formerly Constellation) (Twist et al., 2016), Aldy (Numanagi et al., 2018), and StellarPGx (Twesigomwe et al., 2021). In our analysis with the Stargazer pipeline, *CYP2D6* showed a high degree of polymorphism with a long tail of low-frequency alleles; 33 out of 43 star alleles were identified with a frequency <1%, and 19 out of 33 rare alleles were classified as abnormal function alleles (loss-of-function, decreased function, or increased function alleles, Supplementary Table S1). Several low-frequency alleles have not been reported in the Brazilian population (Rodrigues-Soares et al., 2018), including the loss-of-function rearrangement *CYP2D6**68+*4 (4.5%) and decreased function rearrangement *CYP2D6**36+*10 (frequency <1%). The frequency of the *CYP2D6**5 gene deletion was comparable to the frequency described by Rodrigues-Soares et al. (2020) in 98 Brazilians genotyped for *CYP2D6* polymorphisms. The frequency of *CYP2D6* functional allele multiplication (extensive metabolizers) was 2.5%, corroborating the pattern identified in LLerena et al. (2014) in Latin American populations. SVs were also identified in other 34 of 38 genes analyzed by Stargazer, although they are rare for most of the genes, corroborating Santos et al. (2018), who described novel exonic deletions and duplications in 201 of 208 pharmacogenes analyzed (97%) in the ExAC dataset. Testing the accuracy of SVs identified in PGx-specific pipelines as compared to the best practices available for WGS-based SV and CNV pipeline (Trost et al., 2018) will be followed up in the future.

We verified that 98.0% of the individuals carried at least one high-risk genotype in the nine genes analyzed, and 24.8% of the individuals were predicted to be at a high risk for PGx interactions

for both CYP2C19 and CYP2D6, for example. In terms of analysis of real-world medication data, 42.0% of the individuals used at least one drug with PGx recommendation (PharmGKB level 1A). An important aspect is that our cohort is made up of census-sampled individuals aged 60 years and older with a high rate of polypharmacy, considering that 41.1% of the individuals reported taking five or more medications regularly.

A recent systematic review (O'Shea et al., 2022) indicated that, mainly in a multi-drug scenario, panel-based tests have provided optimistic estimates of long-term cost savings, especially due to a reduction in the number of emergency department (ED) visits and the number of rehospitalizations in patients submitting to PGx tests (Brixner et al., 2016; Elliott et al., 2017). Although an economic analysis is beyond the scope of this study, a previous study conducted in a public hospital in Brazil showed that 14.6% of ED visits were associated with drug-related morbidity, with an estimated annual cost of approximately USD 7.5 million (Freitas et al., 2017).

More research involving economic evaluations of PGx implementation in Brazil is needed and should also include a discussion of the genotyping methodology. Although the idea of using NGS-based techniques in the clinics has been challenged due to the current costs and the complexity involved in the interpretation of results (O'Shea et al., 2022), as we have discussed previously, several advances in terms of techniques have been made in recent years and targeted genotyping tends to continuously be replaced by NGS-based approaches, including WGS (Caspar et al., 2021).

In our real-world in-home medication usage evaluation, *CYP2C19*, *SLCO1B1*, *CYP2D6*, *CYP2C9*, and *CYP4F2* e *VKORC1* were the most important genes from a PGx perspective. *CYP2C19* contributes to the metabolism of a wide range of drugs, including the platelet aggregation inhibitor clopidogrel, proton pump inhibitors, antidepressants, carisoprodol, and diazepam, in addition to endogenous substances, such as melatonin and progesterone (Aquilante et al., 2013; Botton et al., 2021). In our study, *CYP2C19* was the gene with the highest frequency of predicted phenotypes at a potential high risk for an adverse or poor response to medications metabolized by the gene, in agreement with studies in other populations (Biswas, 2021; Luo et al., 2021). Although medications used in the hospital setting were not considered in this analysis, we could envisage the potential clinical impact of *CYP2C19* in terms of real-world medication usage, as drugs that potentially interact with *CYP2C19* had the highest frequency of use among all the gene–drugs pairs analyzed. In total, 25.1% of the cohorts were taking medications with potential *CYP2C19* gene–drug interactions, and 14.5% of the cohorts were high-risk individuals who have actually taken one or more of the drugs at home. The frequency of use of omeprazole was 19.5%, while other drugs were 8.8%, including amitriptyline, sertraline, citalopram, and clopidogrel.

The frequency of individuals at high-risk for *SLCO1B1* gene–drug interaction was 18.7% (with one or two copies of *SLCO1B1**5, *15, or *17), and the real-world in-home medication usage analysis showed that simvastatin was the medication taken with the highest frequency (20.1%) and 3% of the cohort individuals were taking simvastatin with a high risk of adverse drug reactions. The frequency of individuals reporting statin-related skeletal muscle

toxicity, such as myalgias, myopathy, and rhabdomyolysis, is considered low (1–5%), but the high frequency of prescription results in an important absolute number of events (Ramsey et al., 2014).

CYP2C9 is involved in the oxidative metabolism of up to 15–20% of all drugs undergoing phase I metabolisms, such as warfarin, non-steroidal anti-inflammatories, and phenytoin (Lee et al., 2002; Van Booven et al., 2010). In our analysis, ibuprofen, warfarin, and phenytoin were the most commonly used drugs metabolized by CYP2C9. The frequency of decreased function *CYP2C9**2 and *3 varies among populations, and in the Brazilian population, both alleles showed significant differences between self-reported “White” and “Black” individuals (Rodrigues-Soares et al., 2018). *CYP2C9**5, *6, *8, and *11 are decreased function alleles that are found with the highest frequency in individuals of African ancestry⁶. Based on the significant difference in allele frequencies in populations of diverse ancestry, the warfarin guidelines have distinct dosing algorithms in patients who self-identify as African, which includes *CYP2C9**5, *6, *8, *11, vs. non-African. In the SABE sample, the allele frequencies of *CYP2C9**2 and *3 were 11.3% and 5.2%, respectively, and the allele frequencies of *CYP2C9**5, *6, *8, and *11 were <1%. However, the percentage of individuals with at least one copy of *CYP2C9**5, *6, *8, or *11 was 3.92%, and 19.0% of these were self-identified as “White,” 59.5% as “Brown,” and 19.0% as “Black” individuals, corroborating previous studies that demonstrated a tenuous correlation between self-reported “race” and biogeographical ancestry (Pena et al., 2011; Suarez-Kurtz and Parra, 2018).

In addition to *CYP2C9*, warfarin dosing algorithms include *VKORC1**2 (-1639G>A) as a dose-reduction factor, due to the increased sensitivity to warfarin associated with this variant (Johnson et al., 2017). The frequency of individuals with both *CYP2C9* intermediate/poor metabolism phenotypes and *VKORC1* alleles with increased warfarin sensitivity was 14.4% in our cohort. For these individuals, lower doses or an alternative oral anticoagulant might be considered (Johnson et al., 2017). Warfarin dosing algorithms also include the detection of *CYP4F2**3 as optional and only for individuals of non-African ancestry. The carriers of one or two copies of *CYP4F2**3 would require an increase in warfarin dose of 5–10% (Johnson et al., 2017).

The current study has some limitations. First, the initial cohort was composed of 1,343 individuals, but 143 did not have WGS performed because either they did not provide biological samples or the DNA quality did not reach WGS standards. In addition, 29 pairs of individuals had family relationships up to the third degree, which required exclusion of one of the individuals of the pair to avoid rare variant inflation in the population dataset (Naslavsky et al., 2022). Second, the collection of data regarding medication usage was taken in 2010, and we might not have access to all medications taken by individuals, including drugs taken in the hospital setting. The medication intake questionnaire aimed to collect information on drug usage at the moment of interview, and no dosage annotation was obtained. In addition, Stargazer v1.0.7 has a tool for predicting phenotypes that has not yet been systematically validated and was not used in this study. Other pipelines for pharmacogenes

⁶ www.pharmgkb.org/page/cyp2c19RefMaterials

genotyping based on NGS data could be used in the future to validate the results of this same set of data.

Conclusion

Although microarrays have been shown to be a cost-effective tool for the identification of pharmacogenetic variants and dosing adjustments (Reisberg et al., 2019), genome sequencing technologies continue to improve in terms of read length, data analysis, and variant interpretation. Our study used the Stargazer pipeline (Lee et al., 2019a) to call star alleles from a WGS database of 1,171 individuals from the city of Sao Paulo, Brazil. This study illustrates the feasibility of using such techniques on a large scale in the Brazilian population, with the advantage of unraveling complex pharmacogene structures, such as CYP2D6, rare variants, and other CNVs, that have been proved to play an important role in the field of pharmacogenetics and drug response (Santos et al., 2018). The investigation showed that 98.0% of the individuals carried at least one high-risk genotype in 9 PharmGKB evidence level 1A pharmacogenes, and a significant proportion of individuals were at a risk of interaction when the medication report was analyzed. These results call attention to the importance of the systematic adoption of PGx testing for our population and to the use of NGS data to extract pharmacogenetic variants. As important as the introduction of NGS technologies is the validation of such platforms for the call of pharmacogenetic variants. Studies on the economic evaluation of PGx implementation in Brazil are needed and should also include a discussion around the genotyping methodology.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found at <https://abraom.ib.usp.br> (genomic dataset published as a cohort and presented in short variants and frequencies); European Genome-phenome Archive (EGA) hosted by the EBI and the CRG, under EGA Study accession number EGAS00001005052 (individual-level sequence datasets and variant calling datasets).

Ethics statement

The studies involving human participants were reviewed and approved by FSP-USP Institutional Review Board and National Committee of Ethics in Research. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LB-N, MN, EP, and GS-K contributed to the conception and design of the study. MZ, YD, and MN were responsible for

conducting the SABE study and maintaining the database. LB-N and BT performed the bioinformatics analysis. MS and GD assisted with the bioinformatics analysis. SS provided the facility and supervision for bioinformatics analysis. LB-N organized the database and wrote the first draft of the manuscript. All authors contributed to the review of the manuscript, read, and approved the submitted version. An early version of this manuscript was first incorporated into the Ph.D. thesis work of LB-N, which is available online for access at www.teses.usp.br.

Funding

SABE Project and Genomics initiative were funded by Fundação de Amparo à Pesquisa do Estado de São Paulo—FAPESP grants and fellowships (CEPID 2013/08028-1, SABE 2014/50649-6, INCT 2014/50931-3) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq INCT 465355/2014-5 and PQ-2 304746/2022-3). LB-N received a scholarship from Coordination for the Improvement of Higher Education Personnel—Brazilian Ministry of Education (CAPES) and a Globalink Research Award from Mitacs.

Acknowledgments

This research was enabled in part by the support provided by the Digital Research Alliance of Canada (alliancecan.ca).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2023.1178715/full#supplementary-material>

References

- Al-Mahayri, Z. N., Patrinos, G. P., Wattanapokayakit, S., Iemwimangsa, N., Fukunaga, K., Mushihiro, T., et al. (2020). Variation in 100 relevant pharmacogenes among emiratis with insights from understudied populations. *Sci. Rep.* 10, 21310–21315. doi:10.1038/s41598-020-78231-3
- Aquilante, C. L., Niemi, M., Gong, L., Altman, R. B., and Klein, T. E. (2013). PharmGKB summary: Very important pharmacogene information for cytochrome P450, family 2, subfamily C, polypeptide 8. *Pharmacogenet. Genomics* 23, 721–728. doi:10.1097/FPC.0b013e3283653b27
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi:10.1038/nature15393
- Auweru, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* 43, 1–11. doi:10.1002/0471250953.bi1110s43
- Barbarino, J. M., Russ, M. W., and Klein, T. E. (2018). PharmGKB: A worldwide resource for pharmacogenomic information. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 10 (4), e1417. doi:10.1002/wsbm.1417
- Biswas, M. (2021). Global distribution of CYP2C19 risk phenotypes affecting safety and effectiveness of medications. *Pharmacogenomics J.* 21, 190–199. doi:10.1038/s41397-020-00196-3
- Botton, M. R., Whirl-Carrillo, M., Del Tredici, A. L., Sangkuhl, K., Cavallari, L. H., Agúndez, J. A. G., et al. (2021). PharmVar GeneFocus: CYP2C19. *Clin. Pharmacol. Ther.* 109, 352–366. doi:10.1002/cpt.1973
- Brixner, D., Biltaji, E., Bress, A., Unni, S., Ye, X., Mamiya, T., et al. (2016). The effect of pharmacogenetic profiling with a clinical decision support tool on healthcare resource utilization and estimated costs in the elderly exposed to polypharmacy. *J. Med. Econ.* 19, 213–228. doi:10.3111/13696998.2015.1110160
- Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103, 338–348. doi:10.1016/j.ajhg.2018.07.015
- Caspar, S. M., Schneider, T., Stoll, P., Meienberg, J., and Matyas, G. (2021). Potential of whole-genome sequencing-based pharmacogenetic profiling. *Pharmacogenomics* 22, 177–190. doi:10.2217/pgs-2020-0155
- Caudle, K. E., Sangkuhl, K., Whirl-Carrillo, M., Swen, J. J., Haidar, C. E., Klein, T. E., et al. (2020). Standardizing CYP2D6 genotype to phenotype translation: Consensus recommendations from the clinical pharmacogenetics implementation Consortium and Dutch pharmacogenetics working group. *Clin. Transl. Sci.* 13, 116–124. doi:10.1111/CTS.12692
- Cohn, I., Paton, T. A., Marshall, C. R., Basran, R., Stavropoulos, D. J., Ray, P. N., et al. (2017). Genome sequencing as a platform for pharmacogenetic genotyping: A pediatric cohort study. *npj Genomic Med.* 2, 19. doi:10.1038/s41525-017-0021-8
- De Mattia, E., Silvestri, M., Polesel, J., Ecça, F., Mezzalana, S., Scarabel, L., et al. (2022). Rare genetic variant burden in DPYD predicts severe fluoropyrimidine-related toxicity risk. *Biomed. Pharmacother.* 154, 113644. doi:10.1016/j.biopha.2022.113644
- Elliott, L. S., Henderson, J. C., Neradilek, M. B., Moyer, N. A., Ashcraft, K. C., and Thirumaran, R. K. (2017). Clinical impact of pharmacogenetic profiling with a clinical decision support tool in polypharmacy home health patients: A prospective pilot randomized controlled trial. *PLoS One* 12, e0170905. doi:10.1371/journal.pone.0170905
- Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97. doi:10.1038/nrg1767
- Freitas, G. R. M., Tramontina, M. Y., Balbinotto, G., Hughes, D. A., and Heineck, I. (2017). Economic impact of emergency visits due to drug-related morbidity on a Brazilian hospital. *Value Heal. Reg. Issues* 14, 1–8. doi:10.1016/j.vhri.2017.03.003
- Gray, B., Baruteau, A.-E., Antolin, A. A., Pittman, A., Sarganas, G., Molokhia, M., et al. (2022). Rare variation in drug metabolism and long QT genes and the genetic susceptibility to acquired long QT syndrome. *Circ. Genomic Precis. Med.* 15, e003391. doi:10.1161/CIRCGEN.121.003391
- Guan, P., and Sung, W. K. (2016). Structural variation detection using next-generation sequencing data: A comparative technical review. *Methods* 102, 36–49. doi:10.1016/j.ymeth.2016.01.020
- He, Y., Hoskins, J. M., and McLeod, H. L. (2011). Copy number variants in pharmacogenetic genes. *Trends Mol. Med.* 17, 244–251. doi:10.1016/j.molmed.2011.01.007
- Ingelman-Sundberg, M., Mkrchtian, S., Zhou, Y., and Lauschke, V. M. (2018). Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genomics* 12, 26–12. doi:10.1186/s40246-018-0157-3
- Johansson, I., and Ingelman-Sundberg, M. (2009). CNVs of human genes and their implication in pharmacogenetics. *Cytogenet. Genome Res.* 123, 195–204. doi:10.1159/000184709
- Johnson, J., Caudle, K., Gong, L., Whirl-Carrillo, M., Stein, C., Scott, S., et al. (2017). Clinical pharmacogenetics implementation Consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin. Pharmacol. Ther.* 102, 397–404. doi:10.1002/cpt.668
- Lebrão, M. L., and Laurenti, R. (2005). Saúde, bem-estar e envelhecimento: o estudo SABE no município de São Paulo. *Rev. Bras. Epidemiol.* 8, 127–141. doi:10.1590/S1415-790X2005000200005
- Lee, C. R., Goldstein, J. A., and Pieper, J. A. (2002). Cytochrome P450 2C9 polymorphisms: A comprehensive review of the *in-vitro* and human data. *Pharmacogenetics* 12, 251–263. doi:10.1097/00008571-200204000-00010
- Lee, S., Wheeler, M. M., Patterson, K., McGee, S., Dalton, R., Woodahl, E. L., et al. (2019a). Stargazer: A software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. *Genet. Med.* 21, 361–372. doi:10.1038/s41436-018-0054-0
- Lee, S., Wheeler, M. M., Thummel, K. E., and Nickerson, D. A. (2019b). Calling star alleles with Stargazer in 28 pharmacogenes with whole genome sequences. *Clin. Pharmacol. Ther.* 1, 1328–1337. doi:10.1002/cpt.1552
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Llerena, A., Naranjo, M. E. G., Rodrigues-Soares, F., Penas-Lledó, E. M., Fariñas, H., and Tarazona-Santos, E. (2014). Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. *Expert Opin. Drug Metab. Toxicol.* 10, 1569–1583. doi:10.1517/17425255.2014.964204
- Luo, S., Jiang, R., Grzymalski, J. J., Lee, W., Lu, J. T., and Washington, N. L. (2021). Comprehensive allele genotyping in critical pharmacogenes reduces residual clinical risk in diverse populations. *Clin. Pharmacol. Ther.* 110, 759–767. doi:10.1002/cpt.2279
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873. doi:10.1093/bioinformatics/btq559
- Maulekoonphairoj, J., Chamnanphon, M., Khongphatthanayothin, A., Sutjaporn, B., Wandee, P., Poovorawan, Y., et al. (2020). Phenotype prediction and characterization of 25 pharmacogenes in Thais from whole genome sequencing for clinical implementation. *Sci. Rep.* 10, 18969. doi:10.1038/s41598-020-76085-3
- McInnes, G., Lavertu, A., Sangkuhl, K., Klein, T. E., Whirl-Carrillo, M., and Altman, R. B. (2021). Pharmacogenetics at scale: An analysis of the UK biobank. *Clin. Pharmacol. Ther.* 109, 1528–1537. doi:10.1002/cpt.2122
- Naslavsky, M. S., Scliar, M. O., Yamamoto, G. L., Wang, J. Y. T., Zverinova, S., Karp, T., et al. (2022). Whole-genome sequencing of 1,171 elderly admixed individuals from São Paulo, Brazil. *Nat. Commun.* 13, 1004. doi:10.1038/s41467-022-28648-3
- Numanagi, I., Maliki, S., Ford, M., Qin, X., Toji, L., Radovich, M., et al. (2018). Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. *Nat. Commun.* 9, 828. doi:10.1038/s41467-018-03273-1
- O’Shea, J., Ledwidge, M., Gallagher, J., Keenan, C., and Ryan, C. (2022). Pharmacogenetic interventions to improve outcomes in patients with multimorbidity or prescribed polypharmacy: A systematic review. *Pharmacogenomics J.* 22, 89–99. doi:10.1038/s41397-021-00260-6
- Pena, S. D. J., Di Pietro, G., Fuchshuber-Moraes, M., Genro, J. P., Hutz, M. H., Kehdy, F. S. G., et al. (2011). The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 6, e17063. doi:10.1371/journal.pone.0017063
- Popejoy, A. B., and Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature* 538, 161–164. doi:10.1038/538161a
- Raczy, C., Petrovski, R., Saunders, C. T., Chorny, I., Kruglyak, S., Margulies, E. H., et al. (2013). Isaac: Ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* 29, 2041–2043. doi:10.1093/bioinformatics/btt314
- Ramsey, L. B., Johnson, S. G., Caudle, K. E., Haidar, C. E., Voora, D., Wilke, R. A., et al. (2014). The clinical pharmacogenetics implementation Consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. *Clin. Pharmacol. Ther.* 96, 423–428. doi:10.1038/clpt.2014.125
- Reisberg, S., Krebs, K., Lepamets, M., Kals, M., Mägi, R., Metsalu, K., et al. (2019). Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: Challenges and solutions. *Genet. Med.* 21, 1345–1354. doi:10.1038/s41436-018-0337-5
- Relling, M. V., and Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature* 526, 343–350. doi:10.1038/nature15817
- Relling, M. V., Klein, T. E., Gammal, R. S., Whirl-Carrillo, M., Hoffman, J. M., and Caudle, K. E. (2020). The clinical pharmacogenetics implementation Consortium: 10 Years later. *Clin. Pharmacol. Ther.* 107, 171–175. doi:10.1002/cpt.1651
- Rodrigues-Soares, F., Kehdy, F. S. G., Sampaio-Coelho, J., Andrade, P. X. C., Céspedes-Garro, C., Zolini, C., et al. (2018). Genetic structure of pharmacogenetic biomarkers in Brazil inferred from a systematic review and population-based cohorts: A

- RIBEF/EPIGEN-Brazil initiative. *Pharmacogenomics J.* 18, 749–759. doi:10.1038/s41397-018-0015-7
- Rodrigues-Soares, F., Penas-Lledo, E. M., Tarazona-Santos, E., Sosa-Macias, M., Teran, E., Lopez-Lopez, M., et al. (2020). Genomic ancestry, CYP2D6, CYP2C9, and CYP2C19 among Latin Americans. *Clin. Pharmacol. Ther.* 107, 257–268. doi:10.1002/cpt.1598
- Santos, M., Niemi, M., Hiratsuka, M., Kumondai, M., Ingelman-Sundberg, M., Lauschke, V. M., et al. (2018). Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genet. Med.* 20, 622–629. doi:10.1038/gim.2017.156
- Schwarz, U. I., Gulilat, M., and Kim, R. B. (2019). The role of next-generation sequencing in pharmacogenetics and pharmacogenomics. *Cold Spring Harb. Perspect. Med.* 9, a033027. doi:10.1101/cshperspect.a033027
- Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell* 177, 26–31. doi:10.1016/j.cell.2019.02.048
- Suarez-Kurtz, G., and Parra, E. J. (2018). “Population diversity in pharmacogenetics: A Latin American perspective,” in *Pharmacogenetics* (Netherlands: Elsevier Inc.), 133–154. doi:10.1016/bs.apha.2018.02.001
- Suarez-Kurtz, G., and Pena, S. D. (2007). “Pharmacogenetic studies in the Brazilian population,” in *Pharmacogenomics in admixed populations* (United States: Landes Bioscience).
- Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299. doi:10.1038/s41586-021-03205-y
- Trost, B., Walker, S., Wang, Z., Thiruvahindrapuram, B., MacDonald, J. R., Sung, W. W. L., et al. (2018). A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data. *Am. J. Hum. Genet.* 102, 142–155. doi:10.1016/j.ajhg.2017.12.007
- Twesigomwe, D., Drögemöller, B. I., Wright, G. E. B., Siddiqui, A., da Rocha, J., Lombard, Z., et al. (2021). StellarPGx: A nextflow pipeline for calling star alleles in cytochrome P450 genes. *Clin. Pharmacol. Ther.* 110, 741–749. doi:10.1002/cpt.2173
- Twist, G. P., Gaedigk, A., Miller, N. A., Farrow, E. G., Willig, L. K., Dinwiddie, D. L., et al. (2016). Constellation: A tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene, CYP2D6, from whole-genome sequences. *npj Genomic Med.* 1, 15007. doi:10.1038/npjgenmed.2015.7
- Van Booven, D., Marsh, S., McLeod, H., Carrillo, M. W., Sangkuhl, K., Klein, T. E., et al. (2010). Cytochrome P450 2C9-CYP2C9. *Pharmacogenet. Genomics* 20, 277–281. doi:10.1097/FPC.0b013e3283349e84
- Whirl-Carrillo, M., Huddart, R., Gong, L., Sangkuhl, K., Thorn, C. F., Whaley, R., et al. (2021). An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* 110, 563–572. doi:10.1002/cpt.2350