



OPEN ACCESS

EDITED BY
Zhichao Liu,
Boehringer Ingelheim, United States

REVIEWED BY
Sigve Nakken,
Oslo University Hospital, Norway
Tao Qing,
Yale University, United States
Lisa Kalman,
Centers for Disease Control and
Prevention (CDC), United States

*CORRESPONDENCE
Li Guo,
✉ lguo@ipe.ac.cn
Ying Yu,
✉ ying_yu@fudan.edu.cn

SPECIALTY SECTION
This article was submitted to
Pharmacogenetics and
Pharmacogenomics,
a section of the journal
Frontiers in Pharmacology

RECEIVED 31 July 2022
ACCEPTED 16 January 2023
PUBLISHED 26 January 2023

CITATION
Liu Y, Lin Z, Chen Q, Chen Q, Sang L,
Wang Y, Shi L, Guo L and Yu Y (2023),
PAnno: A pharmacogenomics annotation
tool for clinical genomic testing.
Front. Pharmacol. 14:1008330.
doi: 10.3389/fphar.2023.1008330

COPYRIGHT
© 2023 Liu, Lin, Chen, Chen, Sang, Wang,
Shi, Guo and Yu. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

PAnno: A pharmacogenomics annotation tool for clinical genomic testing

Yaqing Liu¹, Zipeng Lin¹, Qingwang Chen¹, Qiaochu Chen¹,
Leqing Sang¹, Yunjin Wang^{2,3}, Leming Shi¹, Li Guo^{4,5*} and Ying Yu^{1*}

¹State Key Laboratory of Genetic Engineering, Human Phenome Institute, School of Life Sciences and Shanghai Cancer Center, Fudan University, Shanghai, China, ²Department of Breast Surgery, Precision Cancer Medicine Center, Key Laboratory of Breast Cancer in Shanghai, Fudan University Shanghai Cancer Center, Shanghai, China, ³Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China, ⁴State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing, China, ⁵School of Chemical Engineering, University of Chinese Academy of Sciences, Beijing, China

Introduction: Next-generation sequencing (NGS) technologies have been widely used in clinical genomic testing for drug response phenotypes. However, the inherent limitations of short reads make accurate inference of diplotypes still challenging, which may reduce the effectiveness of genotype-guided drug therapy.

Methods: An automated Pharmacogenomics Annotation tool (PAnno) was implemented, which reports prescribing recommendations and phenotypes by parsing the germline variant call format (VCF) file from NGS and the population to which the individual belongs.

Results: A ranking model dedicated to inferring diplotypes, developed based on the allele (haplotype) definition and population allele frequency, was introduced in PAnno. The predictive performance was validated in comparison with four similar tools using the consensus diplotype data of the Genetic Testing Reference Materials Coordination Program (GeT-RM) as ground truth. An annotation method was proposed to summarize prescribing recommendations and classify drugs into avoid use, use with caution, and routine use, following the recommendations of the Clinical Pharmacogenetics Implementation Consortium (CPIC), etc. It further predicts phenotypes of specific drugs in terms of toxicity, dosage, efficacy, and metabolism by integrating the high-confidence clinical annotations in the Pharmacogenomics Knowledgebase (PharmGKB). PAnno is available at <https://github.com/PreMedKB/PAnno>.

Discussion: PAnno provides an end-to-end clinical pharmacogenomics decision support solution by resolving, annotating, and reporting germline variants.

KEYWORDS

pharmacogenomics, drug response, genomics, diplotype, haplotype, star allele, drug-genotype, precision medicine

1 Introduction

Next generation sequencing (NGS) based clinical genomic testing has become a powerful strategy in precision therapeutics (Ji et al., 2018; Malone et al., 2020; Tafazoli et al., 2021a). In the area of pharmacogenomics (PGx), it is reflected explicitly in the identification and annotation of germline genetic variants functioning in the absorption, distribution, metabolism, and elimination (ADME) of drugs (Goldstein et al., 2003; Klein et al., 2019).

These variants largely contribute to the inter-individual differences in pharmacokinetics or drug response phenotypes (Rodrigues et al., 2019). Based on the genotype-phenotype association, the efficacy and toxicity of drugs can be revealed with a view to tailoring effective and safe drugs, as well as providing sensible advice on the dosage (Ashley, 2016; Roden et al., 2019).

A pharmacogenetic *allele* (*allele, star allele) or *haplotype* is composed of one or more genetic variants on the same chromosome, and a *diplotype* is formed by a pair of haplotypes of the same gene on homologous chromosomes (Robarge et al., 2007; Lingjun et al., 2014). In addition, *genotype* in this paper refers to variants detected by sequencing platform that will be identified as alleles (haplotypes) and then assigned as diplotypes. The diplotypes bridge the transition from genotype to phenotype and are the basis for precise drug administration. For instance, patients with a homozygous UGT1A1*28 (*28/*28 diplotype) are poor metabolizers, which leads to irreversible toxic effects when taking Belinostat, and require starting with a lower dose (Peer et al., 2016).

Accurate inference of diplotypes is the basis for precise personal genomic interpretation. However, the inability to distinguish parental origin and short-read characteristics of NGS increase the difficulty of the analysis (Browning and Browning, 2011; Snyder et al., 2015; Jin et al., 2018; van der Lee et al., 2020). This is due to the fact that diplotypes often require a joint judgment based on the multiple variants on the maternal and paternal chromosomes. Given the complexity of diplotype phasing, computational tools have been developed, e.g., Astrolabe (Twist et al., 2016), Cyrius (Chen et al., 2021), Aldy (Numanagić et al., 2018), Stargazer (Lee et al., 2019), PharmCAT (Sanguhl et al., 2020), StellarPGx (Twesigomwe et al., 2021), and ImpGX (Klanderma et al., 2022). Nevertheless, there is potential to further improve the accuracy of diplotypes inference for NGS-derived variant call format (VCF) files, which are currently widely used in clinical genomic testing.

Annotation of diplotypes for drug responses phenotypes is another crucial aspect of the PGx clinical application. PharmCAT and ImpGX focus on more than a dozen genes associated with drug metabolizing enzymes and transporter proteins, and translate diplotypes into metabolizer phenotypes on a gene-by-gene basis to correlate dosing guidelines. However, sometimes more than one diplotypes are concerned with a drug, and determining drug dosage based on a single diplotype may be one-sided (Tasa et al., 2019; Shugg et al., 2020). As more and more genotype-phenotype associations have been identified, the number of PGx-related genes and alleles has expanded. For example, the number of genes in the Clinical Pharmacogenetics Implementation Consortium (CPIC) with high-confidence evidence (level A and B) has grown to 33 (CPIC, 2022). In this context, there is a growing need to develop computational tools to more comprehensively report the alleles that may affect drug response.

To address the above issues, we built the Pharmacogenomics Annotation tool (PAnno), an end-to-end automated tool for clinical genomic testing oriented to providing prescribing recommendations and predicting drug response phenotypes. PAnno takes a standard germline VCF file as input to identify PGx-relevant diplotypes. To achieve a more accurate inference, we developed a ranking model considering the differences in allele frequency between populations. In addition, we proposed an annotation method for potential phenotypes that takes into account the one-to-many relationship between drugs and

diplotypes by integrating the clinical annotations of the Pharmacogenomics Knowledgebase (PharmGKB). Finally, PAnno provides a summary of the above results in an HTML report.

2 Materials and methods

2.1 The architecture of PAnno

PAnno consists of two components, diplotype inference, and clinical annotation. By parsing the germline VCF file from NGS and the population to which the individual belongs, PAnno will output the prescribing recommendation, detail of inferred diplotypes, and drug response phenotypes in the form of an HTML report (Figure 1).

2.1.1 Diplotype inference

This component aims to identify PGx alleles on each chromosome and infer the diplotypes from the user-submitted VCF file. PAnno first extracts all allele-related variants from the VCF file based on the pre-collected GRCh38 genomic coordinates of PGx alleles. Next, PAnno classifies the alleles into two categories: single-variant and multi-variant. Single-variant alleles constitute diplotypes that do not involve the judgment of multiple variants and the corresponding genes generally have not yet been standardized by a nomenclature committee, such as rs9923231 for VKORC1. Therefore, PAnno obtains their diplotypes based on the genotype information of the VCF file (e.g., the GT field), and reports according to the locations measured. In addition, PAnno determines the presence and copy number of HLA alleles if the relevant results are provided in the VCF. Multi-variant alleles need to take into account more than one variant on the same chromosome. The most likely diplotype of the gene of interest is inferred by the PAnno ranking model.

2.1.2 Clinical annotation

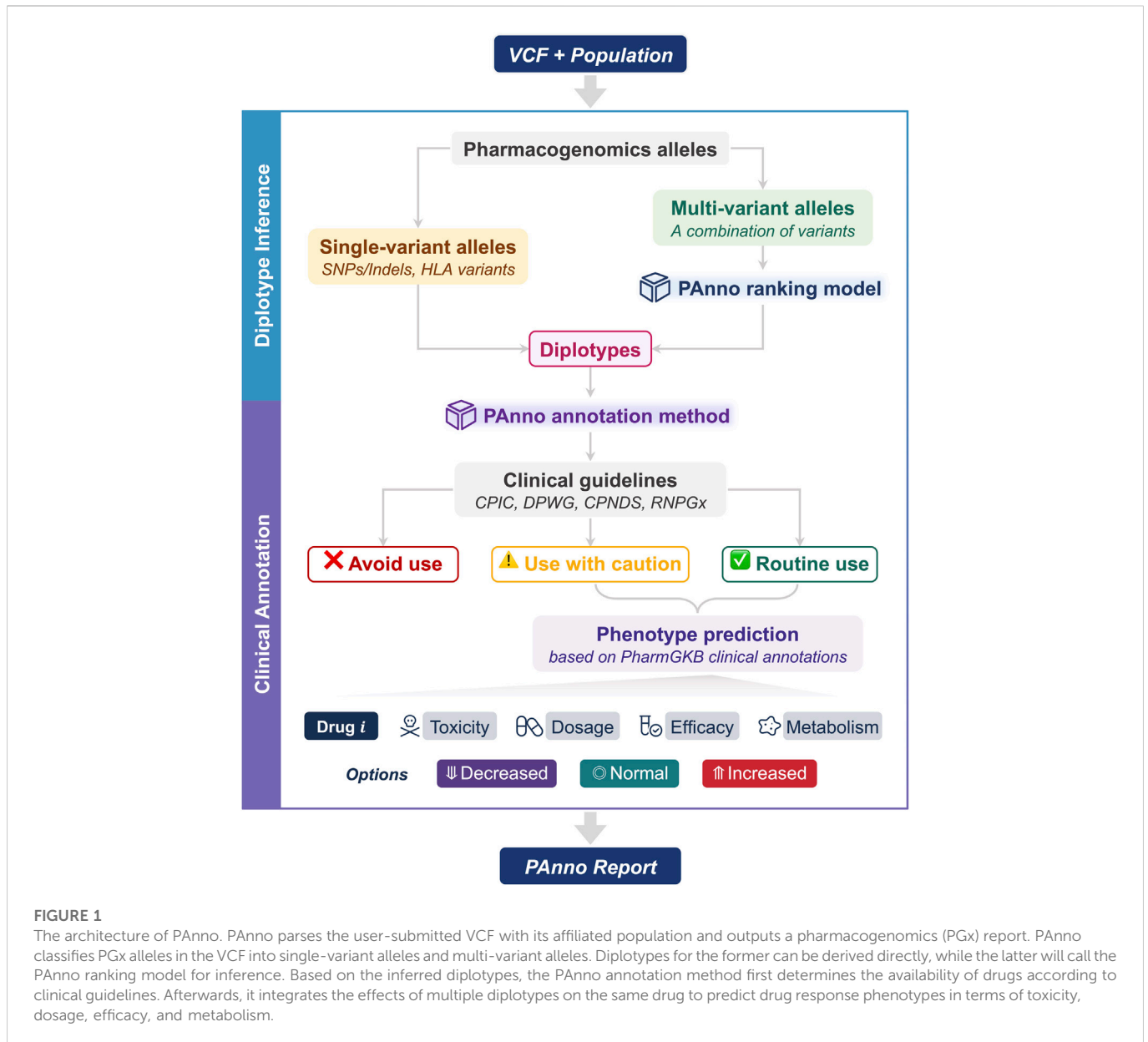
This component aims to translate inferred diplotypes into phenotypes to provide prescribing recommendations and predict drug responses. The PAnno annotation method first annotates diplotypes with dosing guidelines to determine drug availability. For available drugs, clinical annotations of PharmGKB are integrated to predict the possible changes in toxicity, dosage, efficacy, and metabolism of a drug when a patient is routinely taking it.

2.2 Underlying data foundation

2.2.1 Allele definition and population allele frequency

PAnno ranking model relied on the allele definition and population frequency of diplotypes. The PharmGKB (Whirl-Carrillo et al., 2012; Whirl-Carrillo et al., 2021), the Pharmacogene Variation (PharmVar) Consortium (Gaedigk et al., 2021), and the CPIC (Relling and Klein, 2011; Caudle et al., 2014; Relling et al., 2020) are the primary data resource for PAnno to inference diplotypes.

Moreover, PAnno followed the standardized grouping system applied by PharmGKB to divide the population into nine biogeographic groups, namely African American/Afro-Caribbean (AAC), American (AME), Central/South Asian (SAS), East Asian (EAS), European (EUR), Latino (LAT), Near Eastern (NEA),



Oceanian (OCE), and Sub-Saharan African (SSA) (Huddart et al., 2019).

We first obtained the allele definitions and their population frequencies from the PharmGKB website, which provided a further integration of the frequencies compiled by PharmGKB and CPIC based on the published literature reports (Nofziger et al., 2020; Botton et al., 2021; Desta et al., 2021; Sangkuhl et al., 2021; PharmGKB, 2022e; Ramsey et al., 2022; Rodriguez-Antona et al., 2022). For genes for which PharmGKB did not provide population allele frequencies, we estimated the allele frequencies using variant frequencies provided by the 1000 Genomes Project (Siva, 2008). Specifically, we calculated the population frequencies separately for all variants defining each haplotype, and then used the lowest value to represent the frequency for the haplotype. It is worth noting that a null value for population frequency is more likely to result from insufficient data or studies. In contrast, a frequency of zero indicates that the population is

essentially unlikely to have the haplotype. In this case, we assigned a minimal value ($\epsilon = 1e-5$) to the null value.

On this basis, we calculated diplotype-related data based on the following two principles. First, the combination of any two haplotypes of a gene gives all possible diploypes. Second, the combination of all variants involved in two haplotypes equals the definition of the corresponding diplotype. Finally, we annotated the population frequencies to each diplotype of each gene according to the Hardy-Weinberg equilibrium (Hardy, 1908; Weinberg, 1908).

2.2.2 Prescribing recommendations and clinical annotations

Reliable PGx knowledge is the foundation for effective annotation and reporting. We collected the clinical guidelines integrated by PharmGKB and manually reviewed the prescribing recommendations. For given diploypes, we annotated whether the

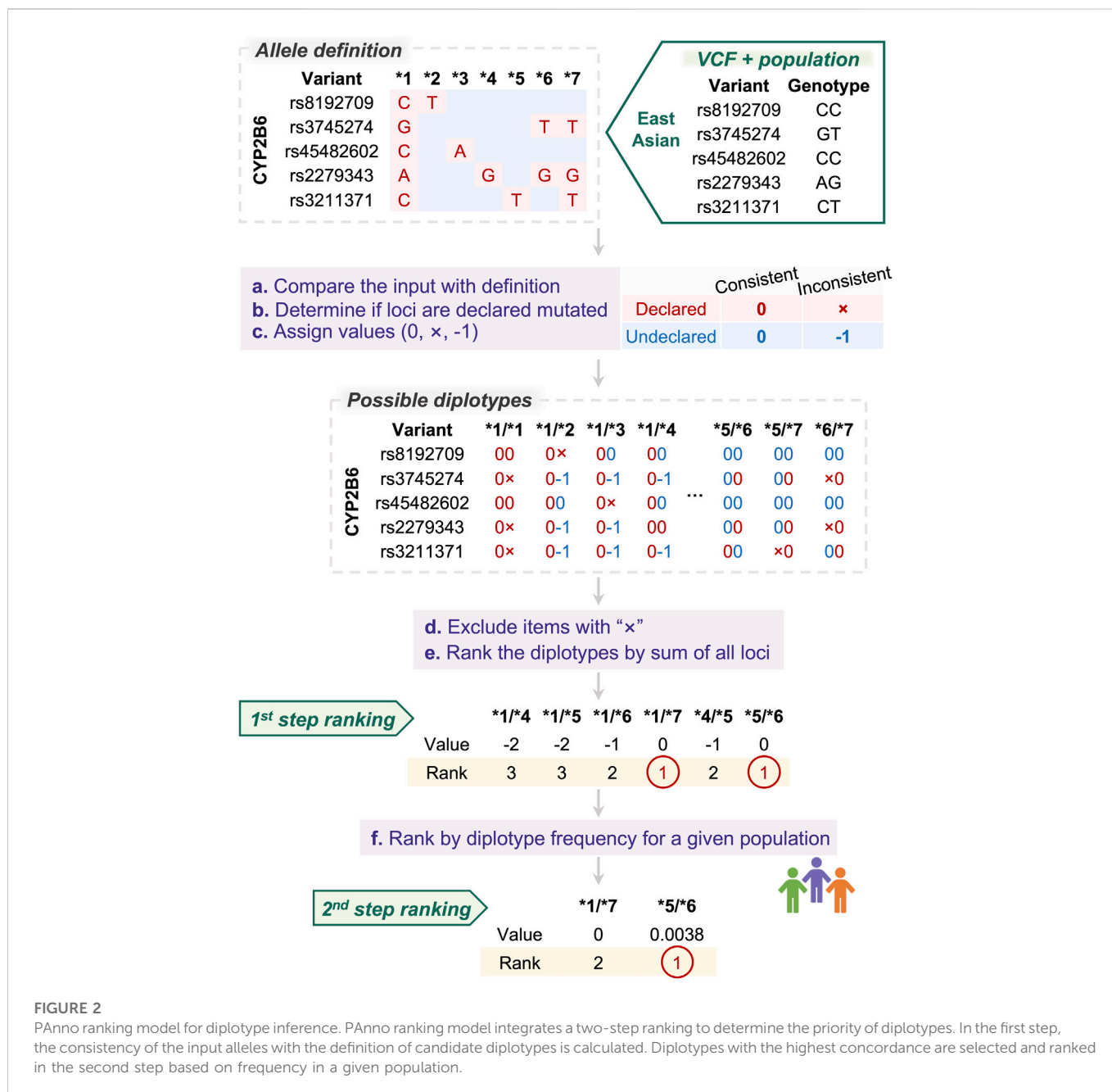


FIGURE 2

PAnno ranking model for diplotype inference. PAnno ranking model integrates a two-step ranking to determine the priority of diplotypes. In the first step, the consistency of the input alleles with the definition of candidate diplotypes is calculated. Diplotypes with the highest concordance are selected and ranked in the second step based on frequency in a given population.

drug in question needs to be avoided, used with caution, or used routinely according to the associated guidelines. The original prescribing information was primarily published by CPIC (Clancy et al., 2014; Martin et al., 2014; Muir et al., 2014; Relling et al., 2014; Birdwell et al., 2015; Hicks et al., 2015; Gammal et al., 2016; Saito et al., 2016; Bell et al., 2017; Hicks et al., 2017; Johnson et al., 2017; Amstutz et al., 2018; Goetz et al., 2018; Phillips et al., 2018; Brown et al., 2019; Desta et al., 2019; Gonsalves et al., 2019; Relling et al., 2019; Theken et al., 2020; Crews et al., 2021; Karnes et al., 2021; Lima et al., 2021; Cooper-DeHoff et al., 2022; Lee et al., 2022; McDermott et al., 2022), the Dutch Pharmacogenetics Working Group (DPWG) (Lunenburg et al., 2020; Brouwer et al., 2021; Matic et al., 2021; DPWG, 2022a; DPWG, 2022b), the Canadian Pharmacogenomics Network for Drug Safety (CPNDS) (Amstutz et al., 2014; Shaw et al., 2015; Aminkeng et al., 2016; Lee et al., 2016; Drögemöller et al., 2019), and the French

National Network for Pharmacogenetics (RNPGx) (Lamoureux and Dufлот, 2017; Picard et al., 2017; Quaranta et al., 2017; Quaranta and Thomas, 2017; Woillard et al., 2017).

In addition, we collected the clinical annotations from PharmGKB which summarize the published evidence for the association between a genetic variant and a drug. Each clinical annotation is assigned a level of evidence including 1A, 1B, 2A, 2B, 3, and 4, from high to low confidence (PharmGKB, 2022d). The variant-drug associations described by 1A, 1B, 2A, and 2B have been demonstrated in multiple trials and even incorporated into dosing guidelines. To improve interpretation reliability and clinical usability, PAnno only included the annotations from these four levels.

A PharmGKB clinical annotation generally corresponds to a drug-genotype pair, which describes in paragraph form what phenotypic response a person would have if they had a certain genotype. We

determined the drug response reflected in the annotation paragraph by regular expressions with manual curation. Inspired by the CIPC standardization project, each annotation was classified into three phenotype terms, namely decreased function, normal function, and increased function (Caudle et al., 2017). For a given clinical description, such as the annotation of rs121434568 GG in <https://www.pharmgkb.org/clinicalAnnotation/981420042>. Our annotated result was that patients taking gefitinib with rs121434568 GG were marked with “increased” in terms of efficacy. To better integrate multiple annotations in practice, these phenotypes were numerically coded as 0.5 (decreased), 1 (normal), and 2 (increased), respectively.

2.3 PAnno ranking model for diplotype inference

PAnno ranking model takes as input genotypes of the positions at which alleles are defined. Candidate diplotypes for each gene will be prioritized by a two-step ranking as described below (Figure 2).

The definitions of alleles that form diplotypes are manually curated by experts. In general, an allele will have variants that are included in the definition of specific positions, which we refer to as “declared.” Loci not included in a haplotype definition are “undeclared.” We presume that the sequences of these undeclared loci are consistent with the wild type.

PAnno first compares the input genotypes with the definition of alleles, and assigns values based on their concordance and whether these loci are declared to be mutated. At declared loci, consistent is marked as “0” and inconsistent as “x”. At undeclared loci, consistent is marked as “0” and inconsistent as “-1.” According to this principle, comparison values will be assigned on all variants of all possible diplotypes.

Next, PAnno excludes diplotypes containing “x”, and calculates the sum of the values of all variants for each diplotype. The higher the total score, i.e., the closer to zero, the more consistent the patient mutation is with the diplotype definition.

Finally, PAnno performs two consecutive steps of ranking. The first ranking is based on the total consistency score. The highest ranked candidates are further ranked according to their frequency in the given population. The diplotypes with the highest rank are the final result of the PAnno ranking model.

2.4 PAnno annotation method for predicting drug response phenotypes

The PAnno annotation method interprets the inferred diplotypes with the associated drugs based on the clinical guidelines of CPIC, DPWG, CPNDS, and RNPGx. Drugs will be classified into three categories: 1) Avoid use. Avoidance of a drug is clearly stated in the prescribing recommendations for the given diplotype. 2) Use with caution. Prescribing changes are recommended for the given diplotype, e.g., dose adjustment and alternative medication. In addition, prescribing recommendations that differ in specific populations or require consideration of multiple diplotypes are included in this category. 3) Routine use. There is no recommended prescribing change for the given diplotype.

For a single drug, there may be multiple associated genotypes, i.e., multiple PharmGKB clinical annotations. To integrate the impact

of multiple genotypes, PAnno calculates the mean of the annotation scores for all associated diplotypes for each clinically available drug. The integrated drug response phenotypes are indicated as decreased (<1), normal (=1), and increased (>1). Notably, drugs judged to be “avoid use” in the guideline-based annotation will not be included in the further interpretation.

Drug response phenotypes predicted by PAnno are further extrapolations beyond metabolizer phenotypes. It is necessary to interpret the PAnno phenotype concerning the corresponding four categories, i.e., toxicity, dosage, efficacy, and metabolism. For example, weak metabolizers tend to obtain higher toxicity after taking the drug at a regular dose, which the PAnno phenotype would be predicted as “decreased metabolism” and “increased toxicity.”

2.5 Test data for GeT-RM samples

The CDC Genetic Testing Reference Materials Coordination Program (GeT-RM) has previously conducted several pharmacogenetic studies to create publicly available genomic DNA reference materials for the field (Pratt et al., 2010; Pratt et al., 2016; Gaedigk et al., 2019; Gaedigk et al., 2022; Pratt et al., 2022). GeT-RM worked with members of the pharmacogenetic testing community to characterize the DNA materials using various techniques such as microarray and sequencing. Consensus diplotypes for a total of 29 PGx genes were developed and used as a reference dataset to assess the quality of PGx analysis. With the development of technology, GeT-RM has updated the results of these samples in earlier studies. Therefore, our evaluation will be based on the latest version of the consensus diplotypes obtained from <https://www.cdc.gov/labquality/GeT-RM/inherited-genetic-diseases-pharmacogenetics/pharmacogenetics.html>.

Among these GeT-RM studies, 137 samples had the most comprehensive PGx genetic assay results. The VCF files for 88 of these samples are freely available from the 1000 Genomes Project at http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20190425_NYGC_GATK/raw_calls_updated/. The germline variant data were re-sequenced and analyzed by the New York Genome Center at 30x coverage (Byrska-Bishop et al., 2021). These samples are from four biogeographic groups, including African American/Afro-Caribbean, East Asian, European, and Latino. The VCF files along with the corresponding population will be used as input to PAnno.

3 Results

3.1 Overview of PAnno

PAnno resolves and annotates PGx alleles based on the patient germline VCF file and the population information to provide prescribing recommendations and predict drug response phenotypes (Figure 1). To facilitate user access, PAnno has been developed as a Python package and a Conda package. The source code, the documentation, and the example reports are available at <https://github.com/PreMedKB/PAnno>.

In the current version, PAnno supports the parsing of diplotypes for 52 genes and the clinical annotation of 100 drugs (Supplementary Table S1). There are 13 genes involving multi-variant defined alleles

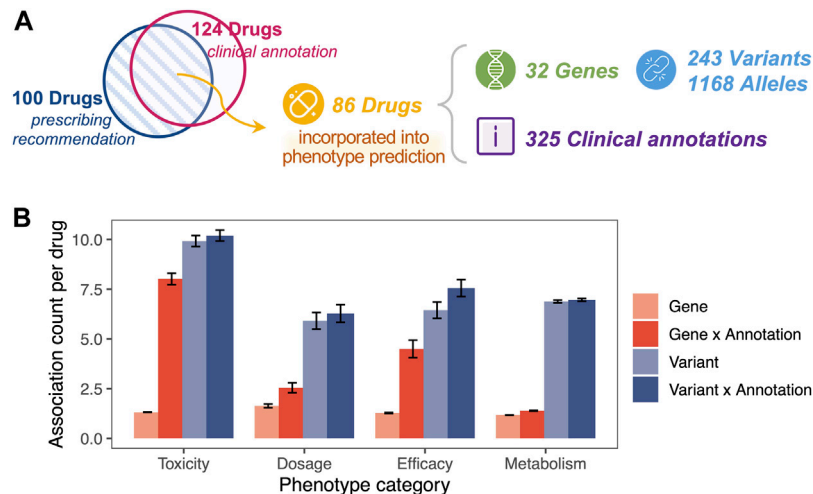


FIGURE 3

One-to-many clinical association for a given drug. **(A)** The PharmGKB clinical annotations used by PAnno covered 86 drugs, 32 genes, 243 variants, and 1168 alleles. The blue circle represents drugs covered in the prescribing recommendations of CPIC, DPWG, CPNDS, and RNPgX guidelines. The pink circle represented drugs involved in clinical annotations with evidence levels of 1A, 1B, 2A, and 2B in PharmGKB. **(B)** Average number of associations for a given drug in relation to genes (light red), combinations of genes and annotations (dark red), variants (light blue), and combinations of variants and annotations (dark blue). Error bars represent \pm standard error of mean. The x-axis indicates the different phenotype categories, including toxicity, dosage, efficacy, and metabolism.

with the PAnno ranking model applied, including CYP2B6, CYP2C19, CYP2C8, CYP2C9, CYP2D6, CYP3A4, CYP3A5, CYP4F2, DPYD, NUDT15, SLCO1B1, TPMT, and UGT1A1. It should be noted that alleles defined on copy number variants (CNV) and structural variants (SV) are not included in PAnno, because CNV and SV variations are not included in typical VCF files. Unincorporated alleles include CYP2B6*29, CYP2B6*30, CYP2C19*36, CYP2C19*37, CYP2D6*5, CYP2D6*13, CYP2D6*61, CYP2D6*63, CYP2D6*68, SLCO1B1*48, and SLCO1B1*49. Additionally, PAnno also includes the assignment of 17 HLA alleles of HLA-A, HLA-B, HLA-C, HLA-DPB1, and HLA-DRB1. The other alleles refer to the single-variant alleles, such as ABCG2, CACNA1S, IFNL3, VKORC1, RYR1, CFTR, etc., which are basically SNVs/Indels.

For clinical annotation, PAnno includes a total of 53 CPIC, 60 DPWG, six CPNDS, and eight RNPgX dosing guidelines, covering 28 genes and 100 drugs. PAnno further collected 325 PharmGKB clinical annotations, of which 282 were at level 1A, 14 at level 1B, 20 at level 2A, and nine at level 2B. These clinical annotations intersected with dosing guidelines for a total of 86 drugs, associated with 32 genes, 243 variants, and 1168 alleles (Figure 3A).

3.2 One-to-many clinical associations for a given drug

One drug was observed with multiple genetic associations in the PharmGKB clinical annotation used by PAnno, even when distinguishing phenotypic categories (Figure 3B). In terms of toxicity, dosage, efficacy, and metabolism, the average number of genes associated with one drug was approximately 1.31, 1.64, 1.28, 1.18, and the average number of variants associated was 9.90, 5.91, 6.28, 6.88, respectively. For a drug-gene pair or a drug-variant pair, there may be more than one clinical

annotation. For example, PharmGKB provides two clinical annotations (PharmGKB ID 1449269910 and 1447673005) to describe the toxicity of warfarin in patients with different genotypes of VKORC1 rs9923231 (PharmGKB, 2022b; PharmGKB, 2022c). The results showed that a significant increase in the average number of gene-annotation combinations occurred for a specific drug, with 8.02, 2.55, 4.5, and 1.39 for toxicity, dosage, efficacy, and metabolism, respectively. Meanwhile, the increase in the number of variant-annotation combinations was not significant, at 10.2, 6.27, 7.06, and 6.96, respectively. The phenomenon of one-to-many associations with a given drug reinforces the importance of the drug-centric nature of the PAnno annotation method.

3.3 Prioritization of indistinguishable diplotypes based on population frequency

3.3.1 Indistinguishable diplotypes for NGS-derived variants

NGS-derived clinical genomics testing data cannot determine exactly which chromosome the variant is located. When two haplotypes are combined, as occurs in a VCF, haplotypes can be inferred based on the allele definitions. When focusing only on the alleles, there may be cases where different diplotypes have the same mutation pattern, and we labeled them as “indistinguishable diplotypes.” We counted the number of haplotypes, diplotypes, and indistinguishable diplotypes for NGS for the 13 genes analyzed by PAnno (Table 1). There were eight genes (CYP2B6, CYP2C9, CYP2C19, CYP2D6, NUDT15, SLCO1B1, TPMT, and UGT1A1) containing indistinguishable diplotypes, which is one of the important reasons for PAnno to set up a further ranking step. The details are included in Supplementary Table S2.

TABLE 1 Number of indistinguishable diplotypes for NGS-derived variants.

Gene	Number of haplotypes	Number of diplotypes	Number of indistinguishable diplotypes
CYP2B6	35	630	20
CYP2C8	18	171	0
CYP2C9	75	2,850	2
CYP2C19	34	595	10
CYP2D6	136	9,316	56
CYP3A4	34	595	0
CYP3A5	6	21	0
CYP4F2	3	6	0
DPYD	83	3,486	0
NUDT15	20	210	2
SLCO1B1	42	903	47
TPMT	46	1,081	2
UGT1A1	10	55	6

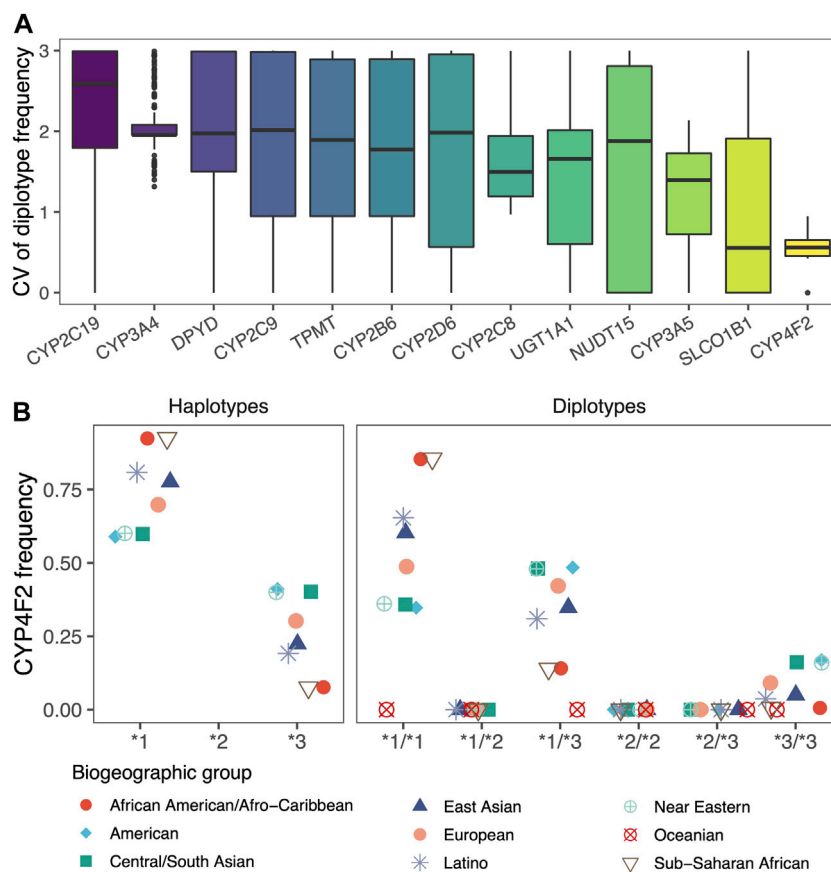


FIGURE 4

Differences in diplotype frequencies across populations. **(A)** Box plots showed the differences in the frequencies of diplotypes in 13 genes across the nine biogeographic groups, which are characterized by coefficients of variation (CV). **(B)** Scatter plot showed the haplotype frequencies of CYP4F2 and the diplotype frequencies calculated according to Hardy-Weinberg equilibrium. Jittering was applied to avoid overlapping.

TABLE 2 Performance of PAnno and similar tools in diplotype inference.

Gene	% PAnno	% PharmCAT	% ImPGX	% Adly	% StellarPGx
CYP2B6	89.53 (77/86)				66.22
CYP2C8	100.00 (88/88)			100.00	100.00
CYP2C9	98.86 (87/88)	100.00	95.70	99.27	100.00
CYP2C19	100.00 (88/88)	100.00	95.70	93.43	97.30
CYP2D6	94.74 (54/57)		94.30	99.08	78.10
CYP3A4	100.00 (40/40)			95.62	100.00
CYP3A5	100.00 (88/88)	100.00	100.00	100.00	100.00
CYP4F2	100.00 (64/64)	80.00	80.30	92.70	95.52
DPYD	100.00 (42/42)	94.92	98.30	57.66	
NUDT15	100.00 (3/3)				
SLCO1B1	70.77 (46/65)	8.47	62.90		
TPMT	100.00 (86/86)	98.31	98.60	98.54	
UGT1A1	56.25 (36/64)	53.33			

3.3.2 Differences in diplotype frequencies across populations

The frequencies of the vast majority of diplotypes in the nine populations had significant differences. For each gene, the average coefficient of variation (CV) of the frequencies in different populations ranged from 0.53 to 2.24. Among them, CYP2C19, CYP3A4, and DPYD had mean CVs greater than 2, indicating that diplotypes of these genes are extremely unlikely to occur in different populations. In addition, we found that CYP2C8 and CYP3A4, which were inferred from the lowest allele frequency based on the 1000 Genomes Project, performed similarly to other genes (Figure 4A).

Even for CYP4F2 with the smallest variability, its haplotypes and diplotypes reflected significant population differences. The population frequency of diplotype obtained by random combinations would be jointly influenced by the two haplotypes (Figure 4B). PAnno takes advantage of the above features to prioritize candidate diplotypes that are indistinguishable based on NGS variants.

3.4 Performance of PAnno in diplotype inference

3.4.1 Concordance between PAnno results and GeT-RM consensus

To evaluate the performance of PAnno in diplotype inference, we applied the ranking model to germline variant data from 88 test samples and calculated the concordance of diplotypes for 14 genes shared in PAnno and GeT-RM (Table 2). Notably, there are differences between the haplotypes included in GeT-RM and PAnno, i.e., CYP2C19*38 is not in GeT-RM, while is considered a wild-type haplotype in PAnno. A detailed comparison of the differences in the allele definition can be found in Supplementary Table S3. Therefore, if the output haplotypes were not included by both GeT-RM and PAnno or did not have a clear conversion relationship, the relevant samples were all excluded.

Overall, the diplotypes predicted by PAnno were highly concordant with the ground truth of GeT-RM. The concordance

rates of diplotypes for CYP2C19, CYP2C8, CYP3A4, CYP3A5, CYP4F2, DPYD, and TPMT were 100%, while for CYP2B6, CYP2C9, CYP2D6, SLCO1B1, and UGT1A1 were 89.53%, 98.86%, 94.74%, 70.77%, and 56.25%, respectively. A detailed description of the consistency concordance is in Supplementary Table S4. The reasons for incomplete concordance can be summarized as two main aspects, i.e., the iteration of the analysis pipeline leading to the VCF files containing variants not found in GeT-RM and the linkage disequilibrium (LD) among haplotypes.

For CYP2B6, one haplotype was inconsistent in seven samples due to the detection of variants defining *4, *5, or *22 in the VCF files, while GeT-RM considered these samples as *1. Two other samples matched both *1/*7 and *5/*6, and due to the high frequency of *5/*6 in the population to which the samples belonged, PAnno did not output *1/*7 as advocated by GeT-RM.

For CYP2C9, an inconsistent sample HG01190 was judged as *2/*61 in GeT-RM, while PAnno inferred it as *1/*61. Similar to the reason for the inconsistency in the seven samples of CYP2B6, the results in the VCF showed a C>T heterozygous transition at rs1799853, which was not satisfying the requirement of *2/*61, i.e., C>T homozygous transition at rs1799853.

For CYP2D6, the inconsistency in the three samples was also due to inconsistent detection of germline variants. GeT-RM considered HG00589 as *1/*21 while PAnno inferred *1/*2 owing to its VCF not covering one of the mutated loci that defined *21 (G>GG at rs72549352). The VCF file of NA18519 contained the variant associated with *106 (C>T at rs28371733) while GeT-RM did not, resulting in a PAnno-inferred diplotype as *29/*106 and GeT-RM as *1/*29. For the diplotype of NA19174, GeT-RM considered *4/*40 while PAnno inferred *4/*58. The main difference between these two diplotypes is that the variant sequence of *40 on rs72549358 is AAAGGGGCG(3) while *58 is AAAGGGGCG(2). The 1000 Genome Project VCF for this sample showed a variant on rs72549358 as AAAGGGGCG(2), which led to the inconsistency.

For SLCO1B1, PAnno inferred the diplotypes of the three inconsistent samples as *1/*14 because the heterozygous variants

defining *14 (A>G at rs2306283 and C>A at rs11045819) were identified. However, GeT-RM considered that no variant occurred and thus judged them as *1/*1. Going back to the original data of GeT-RM, we found that there was a platform where *14 was also inferred, while GeT-RM did not use the results of this platform. The rest of discordance was also since PAnno inferred that the 16 samples carried *37 (A>G at rs2306283) while GeT-RM did not. Fundamentally, it still lies in the difference in variant detection between the GeT-RM project and the VCF files of the 1000 Genomes Project used by PAnno.

For UGT1A1, *60 with *28, *80 with *28, and *80 with *37 have a very high probability of LD (Innocenti et al., 2002; Chen et al., 2014; Gammal et al., 2016). Thus, PAnno extrapolated 12 samples that were labeled *1/*28 in GeT-RM as *60/*80+*28, six samples labeled *28/*28 as *80+*28/*80+*28, eight samples labeled *28/*60 as *28/*80, one sample labeled *28/(**37*) as *80+*28/*80+*37, and one sample labeled *27/*28 as (**27*/*60)/*80+*28. This phenomenon was also found in PharmCAT and was shown to be largely unaffected by the differences in diplotypes with GeT-RM caused by LD on the determination of drug response.

3.4.2 Comparison with other similar tools

We compared the performance of PAnno with four other diplotype phasing tools, including PharmCAT, ImPGX, Adly, and StellarPGx (Table 2). The accuracy of the above tools was derived from their respective publications, implying potential differences in how accuracy is judged. The empty cells in the table do not exactly mean that the software does not support the inference of the corresponding gene, but only that the comparison results with GeT-RM for the tool are not available. For example, PharmCAT supports genes such as CYP2B6 and NUDT15 in subsequent updates. In comparison, PAnno performed excellently in terms of both the number of PGx genes and the accuracy assessed based on GeT-RM consensus. This demonstrates the comparable performance of the PAnno ranking model compared to other excellent algorithms.

Although PAnno's diplotype inference method was developed based on the allele definition, it not only performed exact matching as in PharmCAT and ImPGX but also incorporated population allele frequencies. This allows multiple filtering and ranking of candidate diplotypes. When an exact match to the diplotype is impossible, PAnno will output results that match the allele definition and have a higher probability of occurring in the population. Some variants associated with the haplotype definition were in complex formats or not covered by the current sequencing data. Therefore, they were not resolved successfully in the exact match reflected by the null "symbol.exact", but PAnno still obtained highly consistent results with the GeT-RM consensus. This implies that PAnno may be somewhat tolerant of potential artifacts in variant calling, misfitting of the input format of VCF, etc. Supplementary Table S5 showed the analysis results for 13 genes from 88 test samples at different steps, including diplotypes that matched the definition exactly, diplotypes with the highest concordance with the definition (the first step ranking), diplotypes with the highest population frequency on this basis (the second step ranking).

3.5 PAnno report

PAnno summarizes the analysis results in an HTML report for better understanding by the user. A snapshot of the PAnno report of the 1000 Genomes Project VCF file for NA10859 is used here as an example (Figure 5). The report contains five core sections: 1) Summary

(Figure 5A). Drugs are reported in three categories, namely, avoid use, use with caution, and routine use. This is based on whether clinical guidelines recommend a prescribing change for the given diplotypes. Although the current version of PAnno covers 100 drugs, the report shows only those drugs that the patient's genotype is associated with. Hyperlinks provide direct access to the detailed recommendations in the next section. 2) Prescribing info (Figure 5B). This section is drug-centric and presents all drug-specific genes detected from the patient's genetic profile, along with their corresponding PAnno-inferred diplotypes, CPIC phenotypes, and recommendations in CPIC, DPWG, CPNDS, and RNPgX. 3) Diplotype detail (Figure 5C). The inferred diplotypes for the 13 PGx genes are presented in this section. The detailed information includes the genomic coordinates, the defined and called variants at each position, and potential protein changes. In addition, the genotypes of the single-variant alleles are summarized in a table. 4) Phenotype prediction (Figure 5D). This section summarizes the predicted drug response phenotypes of the PAnno annotation method in terms of toxicity, dosage, efficacy, and metabolism, after integrating the multiple diplotypes associated with a clinically available drug. The predicted phenotypes are indicated as decreased, normal, and increased. 5) Clinical annotation (Figure 5E). The original PharmGKB clinical annotations on which the predicted phenotypes in the previous section are based are listed in a table. In addition, basic information such as sample name, reporting time, biogeographic group, and supplementary notes, will be presented at the beginning and end. PAnno reports of four samples from different populations are available in Supplementary Material S6. Reports for the 88 tested samples are accessible at <https://github.com/PreMedKB/PAnno-analysis>.

4 Discussion

PAnno is an interpretation tool for resolving individual germline genotypes to infer diplotypes and annotate pharmacogenomic knowledge. The frequency of PGx alleles varies across populations, which leads to differences in drug metabolism, transport, etc. (Gaedigk et al., 2017; Zhou et al., 2017; Ahsan et al., 2020; Jithesh et al., 2022). PAnno applied this feature to the PAnno ranking model built for the inference of multi-variant alleles, which facilitated the prioritization of indistinguishable diplotypes from NGS germline variant calling. The validity of PAnno for inferring diplotypes was demonstrated in comparison with the GeT-RM dataset and four similar tools. After resolving PGx-associated diplotypes, a PAnno annotation method was designed to associate clinical dosing guidelines based on the patient's diplotypes and furthermore to integrate the effects of multiple diplotypes associated with a specific drug.

PAnno performs comparably to similar tools in terms of diplotype inference, but the accuracy is still insufficient for some genes involved in complex genomic regions, such as UGT1A1. This is mainly due to the limitation of NGS itself in identifying repetitive sequences, etc., which affects the final haplotype resolution (Mantere et al., 2019; Garg, 2021). However, the detection of CNVs and SVs is critical for the determination of some phenotypes, such as CYP2D6 ultrarapid metabolizers. Since PAnno directly analyzes VCF files, it cannot address the identification of the exact number of CNVs as well as the SVs of large segments. In the current version, some PGx alleles in CYP2B6, CYP2C19, CYP2D6, and SLCO1B1 cannot be analyzed by PAnno. Reviewing the 88 samples in GeT-RM for the above four genes, the diplotypes of CYP2D6 were involved in alleles not contained by PAnno. Specifically, 14 samples involved SVs (*5, *13,

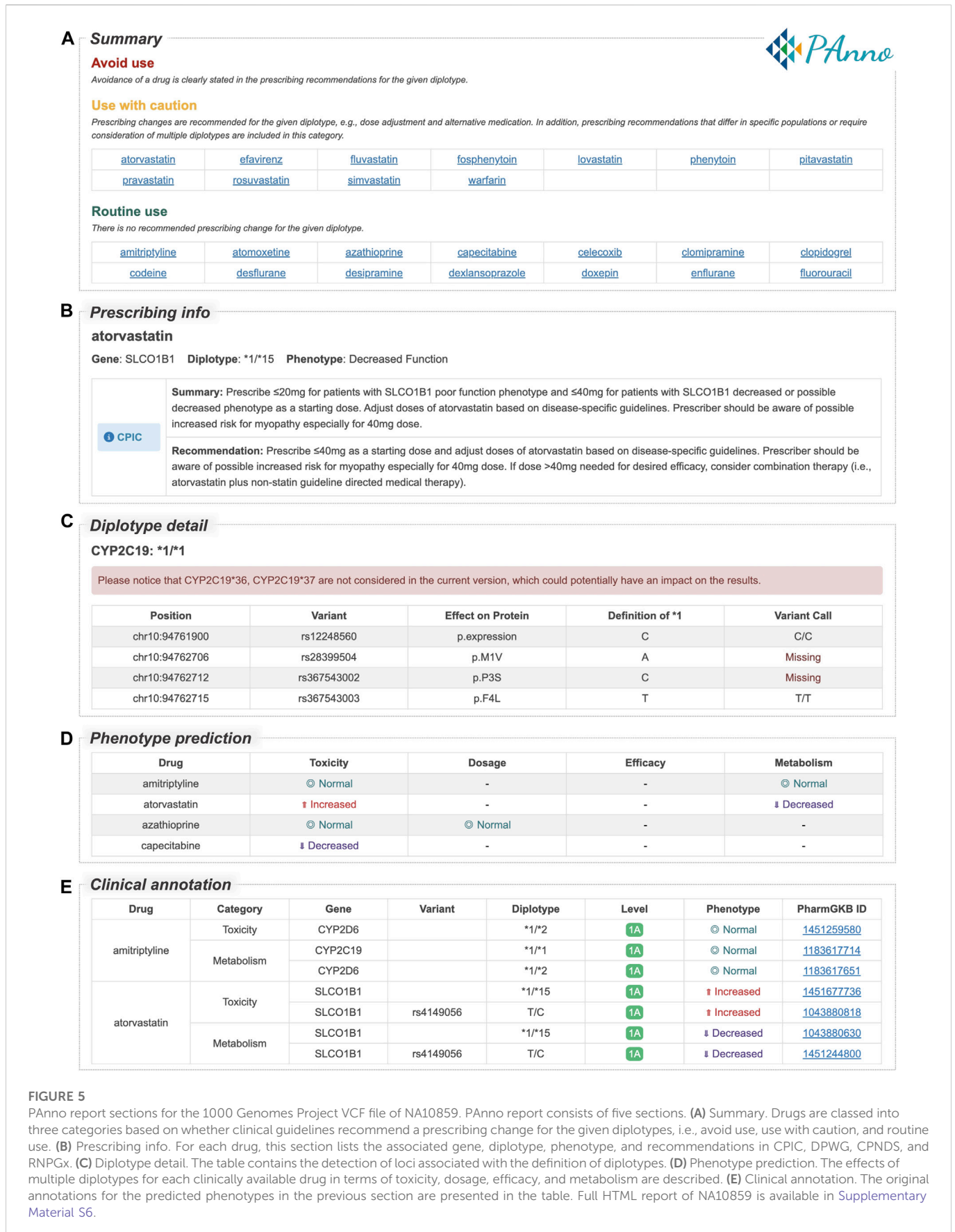


FIGURE 5

Panno report sections for the 1000 Genomes Project VCF file of NA10859. Panno report consists of five sections. (A) Summary. Drugs are classified into three categories based on whether clinical guidelines recommend a prescribing change for the given diplotypes, i.e., avoid use, use with caution, and routine use. (B) Prescribing info. For each drug, this section lists the associated gene, diplotype, phenotype, and recommendations in CPIC, DPWG, CPNDS, and RNPgX. (C) Diplotype detail. The table contains the detection of loci associated with the definition of diplotypes. (D) Phenotype prediction. The effects of multiple diplotypes for each clinically available drug in terms of toxicity, dosage, efficacy, and metabolism are described. (E) Clinical annotation. The original annotations for the predicted phenotypes in the previous section are presented in the table. Full HTML report of NA10859 is available in Supplementary Material S6.

*68), six samples involved tandem arrangements, and eight samples involved CNVs. In total, they account for 35.23% of the overall tested samples, which is a non-negligible proportion (Supplementary Table S4). In this respect, PAnno is deficient compared to Aldy and StellarPGx because they do not rely only on VCF files for the inference of diplotypes. Currently, we have disclosed the non-coverable alleles of CYP2B6, CYP2C19, CYP2D6, and SLCO1B1 in the “diplotype detail” section of the PAnno report, and the compatibility of CNVs and SVs would be considered in the future.

The performance of PAnno is closely related to the quality of the VCF files. Quality here does not only refer to the rigorous filtering of VCFs, but refers broadly to the operational performance throughout the process of data generation and analysis. In this paper, PAnno’s predicted diplotypes using the VCF files of the 1000 Genomes Project have a small number of inconsistencies with the diplotype consensus of GeT-RM. After manual checking, we found many variants that clearly did not match the diplotypes proposed by GeT-RM even in the VCF files. In this case, we are unable to assert the reason for the discrepancies. Therefore, if the germline variants are detected inaccurately, it may raise the risk of inaccurate diplotype inference and even inappropriate clinical recommendations.

Finally, during the clinical annotation step, PAnno does not take into account the effect of disease type. PAnno annotation method only seeks relevant annotations around a specific drug on a specific phenotype category (e.g., toxicity), assigns scores, and integrates them. At this point, some annotations may only be relevant for a specific disease, which may affect the predicted drug response phenotype after integration. In addition, there is a significant amount of manual curation work that limits more precise development. For example, in the annotation of PharmGKB ID 1446846513, neoplasm is not caused by the drug but is the indication for taking anthracyclines (PharmGKB, 2022a). For such clinical annotations specifying indications, complementary disease information contributes to a more accurate pharmacogenomic interpretation. However, this requires laborious manual curation to distinguish indications from adverse effects after drug administration.

Overall, we developed PAnno to provide researchers and clinicians with informative recommendations to aid drug treatment decisions by linking genotypes and drug response phenotypes. The tool has shown great promise in the accuracy and broadness of the analysis of PGx alleles. Furthermore, PGx genes have progressively gone beyond the long-studied range of drug-metabolizing enzymes and transporters. Clinical evidence for nuclear receptors, transcription regulators, and genes potentially related to ADME, etc., is increasing (Arbitrio et al., 2021; Tafazoli et al., 2021b; Lanillos et al., 2022). As the accumulation of associations between different genotypes and drug responses, the integrative strategy for genotype resolution and clinical annotation advocated by PAnno would become more valuable. We believe that the comprehensive interpretation properties of PAnno make it possible to complement the PGx clinical genomic testing.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

YY, LG, and YL conceived the study. YL and ZL developed the diplotype inference model. YL and LQS developed the annotation module and reporting software. YL, QWC, and QCC collected the data used in the development and tested the software. YW contributed to the validation and interpretation of the results. YL drafted the manuscript; LMS, LG, YY, QWC, and QCC reviewed it. All authors contributed to the article and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China (31720103909 and 32170657), the National Key R&D Project of China (2018YFE0201603, and 2021YFF1201305), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01) and the 111 Project (B13016).

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.2023.1008330/full#supplementary-material>

SUPPLEMENTARY TABLE S1

List of PGx genes included in PAnno.

SUPPLEMENTARY TABLE S2

Indistinguishable diplotypes when the parental origin uncertain.

SUPPLEMENTARY TABLE S3

Differences in haplotype definition of PAnno and GeT-RM.

SUPPLEMENTARY TABLE S4

Consistency of PAnno inferred diplotype and GeT-RM consensus.

SUPPLEMENTARY TABLE S5

Diplotypes at different stages of the PAnno ranking model.

SUPPLEMENTARY MATERIAL S6

PAnno reports of four samples from different populations.

References

- Ahsan, T., Urmi, N. J., and Sajib, A. A. (2020). Heterogeneity in the distribution of 159 drug-response related SNPs in world populations and their genetic relatedness. *PLoS ONE* 15, e0228000. doi:10.1371/journal.pone.0228000
- Aminkeng, F., Ross, C. J. D., Rassekh, S. R., Hwang, S., Rieder, M. J., Bhavsar, A. P., et al. (2016). Recommendations for genetic testing to reduce the incidence of anthracycline-induced cardiotoxicity. *Br. J. Clin. Pharmacol.* 82, 683–695. doi:10.1111/bcp.13008
- Amstutz, U., Henricks, L. M., Offer, S. M., Barbarino, J., Schellens, J. H. M., Swen, J. J., et al. (2018). Clinical pharmacogenetics implementation consortium (CPIC) guideline for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing: 2017 update. *Clin. Pharmacol. Ther.* 103, 210–216. doi:10.1002/cpt.911
- Amstutz, U., Shear, N. H., Rieder, M. J., Hwang, S., Fung, V., Nakamura, H., et al. (2014). Recommendations for HLA-B*15:02 and HLA-A*31:01 genetic testing to reduce the risk of carbamazepine-induced hypersensitivity reactions. *Epilepsia* 55, 496–506. doi:10.1111/epi.12564
- Arbitrio, M., Scionti, F., Di Martino, M. T., Caracciolo, D., Pensabene, L., Tassone, P., et al. (2021). Pharmacogenomics biomarker discovery and validation for translation in clinical practice. *Clin. Transl. Sci.* 14, 113–119. doi:10.1111/cts.12869
- Ashley, E. A. (2016). Towards precision medicine. *Nat. Rev. Genet.* 17, 507–522. doi:10.1038/nrg.2016.86
- Bell, G. C., Caudle, K. E., Whirl-Carrillo, M., Gordon, R. J., Hikino, K., Prows, C. A., et al. (2017). Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for CYP2D6 genotype and use of ondansetron and tropisetron. *Clin. Pharmacol. Ther.* 102, 213–218. doi:10.1002/cpt.598
- Birdwell, K. A., Decker, B., Barbarino, J. M., Peterson, J. F., Stein, C. M., Sadee, W., et al. (2015). Clinical pharmacogenetics implementation consortium (CPIC) guidelines for CYP3A5 genotype and tacrolimus dosing. *Clin. Pharmacol. Ther.* 98, 19–24. doi:10.1002/cpt.113
- 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
- Brouwer, J. M. J. L., Nijenhuis, M., Soree, B., Guchelaar, H.-J., Swen, J. J., van Schaik, R. H. N., et al. (2021). Dutch Pharmacogenetics Working Group (DPWG) guideline for the gene-drug interaction between CYP2C19 and CYP2D6 and SSRIs. *Eur. J. Hum. Genet.* 30, 1114–1120. doi:10.1038/s41431-021-01004-7
- Brown, J. T., Bishop, J. R., Sangkuhl, K., Nurmi, E. L., Mueller, D. J., Dinh, J. C., et al. (2019). Clinical pharmacogenetics implementation consortium guideline for cytochrome P450 (CYP)2D6 genotype and atomoxetine therapy. *Clin. Pharmacol. Ther.* 106, 94–102. doi:10.1002/cpt.1409
- Browning, S. R., and Browning, B. L. (2011). Haplotype phasing: Existing methods and new developments. *Nat. Rev. Genet.* 12, 703–714. doi:10.1038/nrg3054
- Byrska-Bishop, M., Evani, U. S., Zhao, X., Basile, A. O., Abel, H. J., Regier, A. A., et al. (2021). High coverage whole genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. bioRxiv, 430068.
- Caudle, K. E., Dunnenberger, H. M., Freimuth, R. R., Peterson, J. F., Burlison, J. D., Whirl-Carrillo, M., et al. (2017). Standardizing terms for clinical pharmacogenetic test results: Consensus terms from the clinical pharmacogenetics implementation consortium (CPIC). *Genet. Med.* 19, 215–223. doi:10.1038/gim.2016.87
- Caudle, K. E., Klein, T. E., Hoffman, J. M., Muller, D. J., Whirl-Carrillo, M., Gong, L., et al. (2014). Incorporation of pharmacogenomics into routine clinical practice: The clinical pharmacogenetics implementation consortium (CPIC) guideline development process. *Curr. Drug Metab.* 15, 209–217. doi:10.2174/1389200215666140130124910
- Chen, X., Shen, F., Gonzalado, N., Malhotra, A., Rogert, C., Taft, R. J., et al. (2021). Cyrius: Accurate CYP2D6 genotyping using whole-genome sequencing data. *Pharmacogenomics J.* 21, 251–261. doi:10.1038/s41397-020-00205-5
- Chen, Z., Su, D., Ai, L., Jiang, X., Wu, C., Xu, Q., et al. (2014). UGT1A1 sequence variants associated with risk of adult hyperbilirubinemia: A quantitative analysis. *Gene* 552, 32–38. doi:10.1016/j.gene.2014.09.009
- Clancy, J. P., Johnson, S. G., Yee, S. W., McDonagh, E. M., Caudle, K. E., Klein, T. E., et al. (2014). Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for ivacaftor therapy in the context of CFTR genotype. *Clin. Pharmacol. Ther.* 95, 592–597. doi:10.1038/clpt.2014.54
- Cooper-DeHoff, R. M., Niemi, M., Ramsey, L. B., Luzum, J. A., Tarkiaainen, E. K., Straka, R. J., et al. (2022). The clinical pharmacogenetics implementation consortium guideline for SLC01B1, ABCG2, and CYP2C9 genotypes and statin-associated musculoskeletal symptoms. *Clin. Pharmacol. Ther.* 111, 1007–1021. doi:10.1002/cpt.2557
- CPIC (2022). *CPIC gene-drug pairs* [online]. Available: <https://cpicpgx.org/genes-drugs/> [Accessed December 01 2022].
- Crews, K. R., Monte, A. A., Huddart, R., Caudle, K. E., Kharasch, E. D., Gaedigk, A., et al. (2021). Clinical pharmacogenetics implementation consortium guideline for CYP2D6, OPRM1, and COMT genotypes and select opioid therapy. *Clin. Pharmacol. Ther.* 110, 888–896. doi:10.1002/cpt.2149
- Destá, Z., El-Boraie, A., Gong, L., Somogyi, A. A., Lausckhe, V. M., Dandara, C., et al. (2021). PharmVar GeneFocus: CYP2B6. *Clin. Pharmacol. Ther.* 110, 82–97. doi:10.1002/cpt.2166
- Destá, Z., Gammal, R. S., Gong, L., Whirl-Carrillo, M., Gaur, A. H., Sukasem, C., et al. (2019). Clinical pharmacogenetics implementation consortium (CPIC) guideline for CYP2B6 and efavirenz-containing antiretroviral therapy. *Clin. Pharmacol. Ther.* 106, 726–733. doi:10.1002/cpt.1477
- DPWG (2022a). *Pharmacogenomic recommendations* [online]. Available: <https://www.g-standaard.nl/risicoanalyse/> [Accessed July 30 2022].
- DPWG (2022b). *Pharmacogenomic recommendations, farmacogenetica-update*. [Online]. Available: <https://www.knmp.nl/> [Accessed July 30 2022].
- Drögemöller, B. I., Wright, G. E. B., Shih, J., Monzon, J. G., Gelmon, K. A., Ross, C. J. D., et al. (2019). CYP2D6 as a treatment decision aid for ER-positive non-metastatic breast cancer patients: A systematic review with accompanying clinical practice guidelines. *Breast Cancer Res. Treat.* 173, 521–532. doi:10.1007/s10549-018-5027-0
- Gaedigk, A., Boone, E. C., Scherer, S. E., Lee, S.-b., Numanagić, I., Sahinalp, C., et al. (2022). CYP2C8, CYP2C9, and CYP2C19 characterization using next-generation sequencing and haplotype analysis: A GeT-RM collaborative project. *J. Mol. Diagnostics* 24, 337–350. doi:10.1016/j.jmoldx.2021.12.011
- Gaedigk, A., Casey, S. T., Whirl-Carrillo, M., Miller, N. A., and Klein, T. E. (2021). Pharmacogene variation consortium: A global resource and repository for pharmacogene variation. *Clin. Pharmacol. Ther.* 110, 542–545. doi:10.1002/cpt.2321
- Gaedigk, A., Sangkuhl, K., Whirl-Carrillo, M., Klein, T., and Leeder, J. S. (2017). Prediction of CYP2D6 phenotype from genotype across world populations. *Genet. Med.* 19, 69–76. doi:10.1038/gim.2016.80
- Gaedigk, A., Turner, A., Everts, R. E., Scott, S. A., Aggarwal, P., Broeckel, U., et al. (2019). Characterization of reference materials for genetic testing of CYP2D6 alleles: A GeT-RM collaborative project. *J. Mol. Diagnostics* 21, 1034–1052. doi:10.1016/j.jmoldx.2019.06.007
- Gammal, R., Court, M., Haidar, C., Iwuchukwu, O., Gaur, A., Alvarellos, M., et al. (2016). Clinical pharmacogenetics implementation consortium (CPIC) guideline for UGT1A1 and atazanavir prescribing. *Clin. Pharmacol. Ther.* 99, 363–369. doi:10.1002/cpt.269
- Garg, S. (2021). Computational methods for chromosome-scale haplotype reconstruction. *Genome Biol.* 22, 101. doi:10.1186/s13059-021-02328-9
- Goetz, M. P., Sangkuhl, K., Guchelaar, H.-J., Schwab, M., Province, M., Whirl-Carrillo, M., et al. (2018). Clinical pharmacogenetics implementation consortium (CPIC) guideline for CYP2D6 and tamoxifen therapy. *Clin. Pharmacol. Ther.* 103, 770–777. doi:10.1002/cpt.1007
- Goldstein, D. B., Tate, S. K., and Sisodiya, S. M. (2003). Pharmacogenetics goes genomic. *Nat. Rev. Genet.* 4, 937–947. doi:10.1038/nrg1229
- Gonsalves, S. G., Dirksen, R. T., Sangkuhl, K., Pulk, R., Alvarellos, M., Vo, T., et al. (2019). Clinical pharmacogenetics implementation consortium (CPIC) guideline for the use of potent volatile anesthetic agents and succinylcholine in the context of RYR1 or CACNA1S genotypes. *Clin. Pharmacol. Ther.* 105, 1338–1344. doi:10.1002/cpt.1319
- Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science* 28, 49–50. doi:10.1126/science.28.706.49
- Hicks, J. K., Bishop, J. R., Sangkuhl, K., Müller, D. J., Ji, Y., Leckband, S. G., et al. (2015). Clinical pharmacogenetics implementation consortium (CPIC) guideline for CYP2D6 and CYP2C19 genotypes and dosing of selective serotonin reuptake inhibitors. *Clin. Pharmacol. Ther.* 98, 127–134. doi:10.1002/cpt.147
- Hicks, J. K., Sangkuhl, K., Swen, J. J., Ellingrod, V. L., Müller, D. J., Shimoda, K., et al. (2017). Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants: 2016 update. *Clin. Pharmacol. Ther.* 102, 37–44. doi:10.1002/cpt.597
- Huddart, R., Fohner, A. E., Whirl-Carrillo, M., Wojcik, G. L., Gignoux, C. R., Popejoy, A. B., et al. (2019). Standardized biogeographic grouping system for annotating populations in pharmacogenetic research. *Clin. Pharmacol. Ther.* 105, 1256–1262. doi:10.1002/cpt.1322
- Innocenti, F., Grimsley, C., Das, S., Ramírez, J., Cheng, C., Kuttub-Boulos, H., et al. (2002). Haplotype structure of the UDP-glucuronosyltransferase 1A1 promoter in different ethnic groups. *Pharmacogenetics Genomics* 12, 725–733. doi:10.1097/00008571-200212000-00006
- Ji, Y., Si, Y., McMillin, G. A., and Lyon, E. (2018). Clinical pharmacogenomics testing in the era of next generation sequencing: Challenges and opportunities for precision medicine. *Expert Rev. Mol. Diagnostics* 18, 411–421. doi:10.1080/14737159.2018.1461561
- Jin, Y., Wang, J., Bachtiar, M., Chong, S. S., and Lee, C. G. L. (2018). Architecture of polymorphisms in the human genome reveals functionally important and positively selected variants in immune response and drug transporter genes. *Hum. Genomics* 12, 43. doi:10.1186/s40246-018-0175-1
- Jithesh, P. V., Abuhaliqa, M., Syed, N., Ahmed, I., El Anbari, M., Bastaki, K., et al. (2022). A population study of clinically actionable genetic variation affecting drug response from the Middle East. *npj Genomic Med.* 7, 10. doi:10.1038/s41525-022-00281-5
- Johnson, J. A., Caudle, K. E., Gong, L., Whirl-Carrillo, M., Stein, C. M., Scott, S. A., 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

- Karnes, J. H., Rettie, A. E., Somogyi, A. A., Huddart, R., Fohner, A. E., Formea, C. M., et al. (2021). Clinical pharmacogenetics implementation consortium (CPIC) guideline for CYP2C9 and HLA-B genotypes and phenytoin dosing: 2020 update. *Clin. Pharmacol. Ther.* 109, 302–309. doi:10.1002/cpt.2008
- Klanderma, B. J., Koch, C., Machini, K., Parpattedar, S. S., Bandyadka, S., Lin, C.-F., et al. (2022). Automated pharmacogenomic reports for clinical genome sequencing. *J. Mol. Diagnostics* 24, 205–218. doi:10.1016/j.jmoldx.2021.12.001
- Klein, K., Tremmel, R., Winter, S., Fehr, S., Battke, F., Scheurenbrand, T., et al. (2019). A new panel-based next-generation sequencing method for ADME genes reveals novel associations of common and rare variants with expression in a human liver cohort. *Front. Genet.* 10, 7. doi:10.3389/fgene.2019.00007
- Lamoureux, F., and Duflo, T. French Network of Pharmacogenetics RNPX (2017). Pharmacogenetics in cardiovascular diseases: State of the art and implementation recommendations of the French national Network of pharmacogenetics (RNPX). *Therapies* 72, 257–267. doi:10.1016/j.therap.2016.09.017
- Lanillos, J., Carcajona, M., Maietta, P., Alvarez, S., and Rodriguez-Antona, C. (2022). Clinical pharmacogenetic analysis in 5,001 individuals with diagnostic Exome Sequencing data. *npj Genomic Med.* 7, 12. doi:10.1038/s41525-022-00283-3
- Lee, C. R., Luzum, J. A., Sangkuhl, K., Gammal, R. S., Sabatine, M. S., Stein, C. M., et al. (2022). Clinical pharmacogenetics implementation consortium guideline for CYP2C19 genotype and clopidogrel therapy: 2022 update. *Clin. Pharmacol. Ther.*
- Lee, J. W., Pussegoda, K., Rassekh, S. R., Monzon, J. G., Liu, G., Hwang, S., et al. (2016). Clinical practice recommendations for the management and prevention of cisplatin-induced hearing loss using pharmacogenetic markers. *Ther. Drug Monit.* 38, 423–431. doi:10.1097/FTD.0000000000000298
- Lee, S.-B., Wheeler, M. M., Patterson, K., McGee, S., Dalton, R., Woodahl, E. L., et al. (2019). 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
- Lima, J. J., Thomas, C. D., Barbarino, J., Desta, Z., Van Driest, S. L., El Rouby, N., et al. (2021). Clinical pharmacogenetics implementation consortium (CPIC) guideline for CYP2C19 and proton pump inhibitor dosing. *Clin. Pharmacol. Ther.* 109, 1417–1423. doi:10.1002/cpt.2015
- Lingjun, Z., Kesheng, W., and Xingguang, L. (2014). Use of diplotypes - matched haplotype pairs from homologous chromosomes - in gene-disease association studies. *Shanghai Arch. Psychiatry* 26, 165–170. doi:10.3969/j.issn.1002-0829.2014.03.009
- Lunenburg, C. A. T. C., van der Wouden, C. H., Nijenhuis, M., Crommentuijn-van Rheden, M. H., de Boer-Veger, N. J., Buunk, A. M., et al. (2020). Dutch Pharmacogenetics Working Group (DPWG) guideline for the gene-drug interaction of DPYD and fluoropyrimidines. *Eur. J. Hum. Genet.* 28, 508–517. doi:10.1038/s41431-019-0540-0
- Malone, E. R., Oliva, M., Sabatini, P. J. B., Stockley, T. L., and Siu, L. L. (2020). Molecular profiling for precision cancer therapies. *Genome Med.* 12, 8. doi:10.1186/s13073-019-0703-1
- Mantere, T., Kersten, S., and Hoischen, A. (2019). Long-read sequencing emerging in medical genetics. *Front. Genet.* 10, 426. doi:10.3389/fgene.2019.00426
- Martin, M. A., Hoffman, J. M., Freimuth, R. R., Klein, T. E., Dong, B. J., Pirmohamed, M., et al. (2014). Clinical pharmacogenetics implementation consortium guidelines for HLA-B genotype and abacavir dosing: 2014 update. *Clin. Pharmacol. Ther.* 95, 499–500. doi:10.1038/clpt.2014.38
- Matic, M., Nijenhuis, M., Soree, B., de Boer-Veger, N. J., Buunk, A.-M., Houwink, E. J. F., et al. (2021). Dutch Pharmacogenetics Working Group (DPWG) guideline for the gene-drug interaction between CYP2D6 and opioids (codeine, tramadol and oxycodone). *Eur. J. Hum. Genet.* 30, 1105–1113. doi:10.1038/s41431-021-00920-y
- McDermott, J. H., Wolf, J., Hoshitsuki, K., Huddart, R., Caudle, K. E., Whirl-Carrillo, M., et al. (2022). Clinical pharmacogenetics implementation consortium guideline for the use of aminoglycosides based on MT-RNR1 genotype. *Clin. Pharmacol. Ther.* 111, 366–372. doi:10.1002/cpt.2309
- Muir, A. J., Gong, L., Johnson, S. G., Lee, M. T. M., Williams, M. S., Klein, T. E., et al. (2014). Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for IFNL3 (IL28B) genotype and PEG interferon- α -based regimens. *Clin. Pharmacol. Ther.* 95, 141–146. doi:10.1038/clpt.2013.203
- Nofziger, C., Turner, A. J., Sangkuhl, K., Whirl-Carrillo, M., Agúndez, J. A. G., Black, J. L., et al. (2020). PharmVar GeneFocus: CYP2D6. *Clin. Pharmacol. Ther.* 107, 154–170. doi:10.1002/cpt.1643
- Numanagic, I., Malikić, 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
- Peer, C. J., Goey, A. K. L., Sissung, T. M., Erlich, S., Lee, M.-J., Tomita, Y., et al. (2016). UGT1A1 genotype-dependent dose adjustment of belinostat in patients with advanced cancers using population pharmacokinetic modeling and simulation. *J. Clin. Pharmacol.* 56, 450–460. doi:10.1002/jcph.627
- PharmGKB (2022a). Clinical Annotation for rs2229774 (RARG); anthracyclines and related substances; cardiotoxicity and Neoplasms (level 3 Toxicity) [Online]. Available: <https://www.pharmgkb.org/clinicalAnnotation/1446846513> [Accessed July 15 2022].
- PharmGKB (2022b). Clinical annotation for rs9923231 (VKORC1); warfarin; hemorrhage (level 2A toxicity) [online]. Available: <https://www.pharmgkb.org/clinicalAnnotation/1449269910> [Accessed November 27 2022].
- PharmGKB (2022c). Clinical annotation for rs9923231 (VKORC1); warfarin; over-anticoagulation (level 1B toxicity) [online]. Available: <https://www.pharmgkb.org/clinicalAnnotation/1447673005> [Accessed November 27 2022].
- PharmGKB (2022d). Clinical annotation levels of evidence [online]. Available: <https://www.pharmgkb.org/page/clinAnnLevels> [Accessed July 30 2022].
- PharmGKB (2022e). PGx gene-specific information tables [Online]. Available: <https://www.pharmgkb.org/page/pgxGeneRef> [Accessed July 30 2022].
- Phillips, E. J., Sukasem, C., Whirl-Carrillo, M., Müller, D. J., Dunnenberger, H. M., Chantrata, W., et al. (2018). Clinical pharmacogenetics implementation consortium guideline for HLA genotype and use of carbamazepine and oxcarbazepine: 2017 update. *Clin. Pharmacol. Ther.* 103, 574–581. doi:10.1002/cpt.1004
- Picard, N., Boyer, J. C., Etienne-Grimaldi, M. C., Barin-Le Guellec, C., Thomas, F., Loriot, M. A., et al. (2017). Pharmacogenetics-based personalized therapy: Levels of evidence and recommendations from the French Network of Pharmacogenetics (RNPX). *Therapies* 72, 185–192. doi:10.1016/j.therap.2016.09.014
- Pratt, V. M., Everts, R. E., Aggarwal, P., Beyer, B. N., Broeckel, U., Epstein-Baak, R., et al. (2016). Characterization of 137 genomic DNA reference materials for 28 pharmacogenetic genes: A GeT-RM collaborative project. *J. Mol. Diagnostics* 18, 109–123. doi:10.1016/j.jmoldx.2015.08.005
- Pratt, V. M., Wang, W. Y., Boone, E. C., Broeckel, U., Cody, N., Edelmann, L., et al. (2022). Characterization of reference materials for TPMT and NUDT15: A GeT-RM collaborative project. *J. Mol. Diagnostics* 24, 1079–1088. doi:10.1016/j.jmoldx.2022.06.008
- Pratt, V. M., Zehnbauser, B., Wilson, J. A., Baak, R., Babic, N., Bettinotti, M., et al. (2010). Characterization of 107 genomic DNA reference materials for CYP2D6, CYP2C19, CYP2C9, VKORC1, and UGT1A1: A GeT-RM and association for molecular pathology collaborative project. *J. Mol. Diagnostics* 12, 835–846. doi:10.2353/jmoldx.2010.100090
- Quaranta, S., Dupouey, J., Colle, R., and Verstuyft, C. (2017). Pharmacogenetics of antidepressant drugs: State of the art and clinical implementation - recommendations from the French National Network of Pharmacogenetics. *Therapies* 72, 311–318. doi:10.1016/j.therap.2016.09.018
- Quaranta, S., and Thomas, F. (2017). Pharmacogenetics of anti-cancer drugs: State of the art and implementation - recommendations of the French National Network of Pharmacogenetics. *Therapies* 72, 205–215. doi:10.1016/j.therap.2017.01.005
- Ramsey, L. B., Gong, L., Lee, S.-b., Wagner, J. B., Zhou, X., Sangkuhl, K., et al. (2022). PharmVar GeneFocus: SLCO1B1. *Clin. Pharmacol. Ther.*
- Relling, M., and Klein, T. (2011). CPIC: Clinical pharmacogenetics implementation consortium of the pharmacogenetics research Network. *Clin. Pharmacol. Ther.* 89, 464–467. doi:10.1038/clpt.2010.279
- 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
- Relling, M. V., McDonagh, E. M., Chang, T., Caudle, K. E., McLeod, H. L., Haidar, C. E., et al. (2014). Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for rasburicase therapy in the context of G6PD deficiency genotype. *Clin. Pharmacol. Ther.* 96, 169–174. doi:10.1038/clpt.2014.97
- Relling, M. V., Schwab, M., Whirl-Carrillo, M., Suarez-Kurtz, G., Pui, C.-H., Stein, C. M., et al. (2019). Clinical pharmacogenetics implementation consortium guideline for thiopurine dosing based on TPMT and NUDT15 genotypes: 2018 update. *Clin. Pharmacol. Ther.* 105, 1095–1105. doi:10.1002/cpt.1304
- Robarge, J. D., Li, L., Desta, Z., Nguyen, A., and Flockhart, D. A. (2007). The star-allele nomenclature: Retooling for translational genomics. *Clin. Pharmacol. Ther.* 82, 244–248. doi:10.1038/sj.clpt.6100284
- Roden, D. M., McLeod, H. L., Relling, M. V., Williams, M. S., Mensah, G. A., Peterson, J. F., et al. (2019). Pharmacogenomics. *Lancet* 394, 521–532. doi:10.1016/S0140-6736(19)31276-0
- Rodrigues, J. C. G., Fernandes, M. R., Guerreiro, J. F., da Silva, A. L. d. C., Ribeiro-dos-Santos, A., Santos, S., et al. (2019). Polymorphisms of ADME-related genes and their implications for drug safety and efficacy in Amazonian Amerindians. *Sci. Rep.* 9, 7201. doi:10.1038/s41598-019-43610-y
- Rodriguez-Antona, C., Savio, J. L., Lauschke, V. M., Sangkuhl, K., Drögemöller, B. I., Wang, D., et al. (2022). PharmVar GeneFocus: CYP3A5. *Clin. Pharmacol. Ther.*
- Saito, Y., Stamp, L. K., Caudle, K. E., Hershfield, M. S., McDonagh, E. M., Callaghan, J. T., et al. (2016). Clinical pharmacogenetics implementation consortium (CPIC) guidelines for human leukocyte antigen B (HLA-B) genotype and allopurinol dosing: 2015 update. *Clin. Pharmacol. Ther.* 99, 36–37. doi:10.1002/cpt.161
- Sangkuhl, K., Claudio-Campos, K., Cavallari, L. H., Agúndez, J. A. G., Whirl-Carrillo, M., Duong, J., et al. (2021). PharmVar GeneFocus: CYP2C9. *Clin. Pharmacol. Ther.* 110, 662–676. doi:10.1002/cpt.2333
- Sangkuhl, K., Whirl-Carrillo, M., Whaley, R. M., Woon, M., Lavertu, A., Altman, R. B., et al. (2020). Pharmacogenomics clinical annotation tool (PharmCAT). *Clin. Pharmacol. Ther.* 107, 203–210. doi:10.1002/cpt.1568
- Shaw, K., Amstutz, U., Kim, R. B., Lesko, L. J., Turgeon, J., Michaud, V., et al. (2015). Clinical practice recommendations on genetic testing of CYP2C9 and VKORC1 variants in warfarin therapy. *Ther. Drug Monit.* 37, 428–436. doi:10.1097/FTD.0000000000000192
- Shugg, T., Pasternak, A. L., London, B., and Luzum, J. A. (2020). Prevalence and types of inconsistencies in clinical pharmacogenetic recommendations among major U.S. sources. *npj Genomic Med.* 5, 48. doi:10.1038/s41525-020-00156-7
- Siva, N. (2008). 1000 Genomes project. *Nat. Biotechnol.* 26, 256. doi:10.1038/nbt0308-256b

- Snyder, M. W., Adey, A., Kitzman, J. O., and Shendure, J. (2015). Haplotype-resolved genome sequencing: Experimental methods and applications. *Nat. Rev. Genet.* 16, 344–358. doi:10.1038/nrg3903
- Tafazoli, A., Guchelaar, H.-J., Milyk, W., Kretowski, A. J., and Swen, J. J. (2021a). Applying next-generation sequencing platforms for pharmacogenomic testing in clinical practice. *Front. Pharmacol.* 12, 693453. doi:10.3389/fphar.2021.693453
- Tafazoli, A., Wawrusiewicz-Kurylonek, N., Posmyk, R., and Milyk, W. (2021b). Pharmacogenomics, how to deal with different types of variants in next generation sequencing data in the personalized medicine area. *J. Clin. Med.* 10, 34. doi:10.3390/jcm10010034
- Tasa, T., Krebs, K., Kals, M., Mägi, R., Lauschke, V. M., Haller, T., et al. (2019). Genetic variation in the Estonian population: Pharmacogenomics study of adverse drug effects using electronic health records. *Eur. J. Hum. Genet.* 27, 442–454. doi:10.1038/s41431-018-0300-6
- Theken, K. N., Lee, C. R., Gong, L., Caudle, K. E., Formea, C. M., Gaedigk, A., et al. (2020). Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2C9 and nonsteroidal anti-inflammatory drugs. *Clin. Pharmacol. Ther.* 108, 191–200. doi:10.1002/cpt.1830
- 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 der Lee, M., Kriek, M., Guchelaar, H.-J., and Swen, J. J. (2020). Technologies for pharmacogenomics: A review. *Genes* 11, 1456. doi:10.3390/genes11121456
- Weinberg, W. (1908). On the demonstration of heredity in man. *Papers on human genetics*.
- 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
- Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., et al. (2012). Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* 92, 414–417. doi:10.1038/clpt.2012.96
- Woillard, J. B., Chouchana, L., Picard, N., and Lorient, M. A. French Network of Pharmacogenetics RNPgX (2017). Pharmacogenetics of immunosuppressants: State of the art and clinical implementation - recommendations from the French National Network of Pharmacogenetics (RNPgX). *Therapies* 72, 285–299. doi:10.1016/j.therap.2016.09.016
- Zhou, Y., Ingelman-Sundberg, M., and Lauschke, V. M. (2017). Worldwide distribution of cytochrome P450 alleles: A meta-analysis of population-scale sequencing projects. *Clin. Pharmacol. Ther.* 102, 688–700. doi:10.1002/cpt.690