Supplemental Note:

Details to the Methods

Section A3) Allele definitions

Allele definitions in the CPIC/PharmGKB tables are derived from the Pharmacogene Variation Consortium (PharmVar; https://pharmvar.org), the TPMT Nomenclature Committee (https://www.imh.liu.se/tpmtalleles?l=en), the UGT Nomenclature Committee (http://prime.vetmed.wsu.edu/resources/udp-glucuronsyltransferase-homepage), or from publications evaluated by the guideline authors for SLCO1B1.

CPIC recommendations are based on star allele diplotypes for *CYP2C19*, *CYP2C9*, *CYP3A5*, *CYP4F2*, *SLCO1B1*, *TPMT* and *UGT1A1*. The tables for these genes contain rows for each star allele. The first row is the "reference" row and represents the *1 allele. Variants within the individual alleles are represented as columns in the file. Nucleotide and protein sequence identifiers, as well as dbSNP rsIDs when available, are provided for each variant. CPIC recommendations for *CFTR, DPYD, IFNL3* and *VKORC1* are based on variant genotypes. The tables for these genes contain rows for each variant, and columns for dbSNP rsIDs and nucleotide and protein sequence identifiers.

Section B2) SLCO1B1 and UGT1A1 rule-based systems in the Reporter

If *SLCO1B1* genotypes are returned by the Named Allele Matcher, that information is used by the Reporter. If, however, the Named Allele Matcher is not able to find a

genotype match, the rs4149056 genotype is used if present in the VCF file. The C allele at rs4149056 is the only variant present in *SLCO1B1*5* and is also found in combination with other variants in *15 and *17. The CPIC guideline for simvastatin includes recommendations for the rs4149056 genotype alone, as well as the star alleles (1).

For phased data, *UGT1A1* matches are displayed as a genotype listing all variants per chromosome based on the VCF file input. For unphased data, all allele combinations the Named Allele Matcher determines based on the input VCF file are listed, which can result in repetition of alleles (*e.g.*, in the case of homozygous alleles). The resulting allele functionality and metabolizer phenotype is determined by *UGT1A1* specific logic. The algorithm accounts for the November 2017 CPIC website update to the UGT1A1/atazanavir guideline which states that when *80, *28 and *37 are all assayed and only *80 is found, there are not enough clinical data to predict metabolizer status with certainty. However, in cases where *80 is assayed but *28 and *37 are not assayed, metabolizer status may be inferred from *80 alone. Therefore, the *UGT1A1* PharmCAT allele definition file does not contain a row for *80 alone, but instead contains rows for (*80 + *28) and (*80 + *37). When PharmCAT detects *80 but information about *28 and *37 is missing in the VCF file, the matcher will provide both genotypes as possibilities.

Section C) Output – report

The PharmCAT report is divided in 4 sections, 1) Genotype Summary; 2) CPIC Recommendation; 3) Allele Call Details; 4) Disclaimers, which are described in the supplemental material.

1. Genotype Summary. The first section summarizes the relevant drugs and genotype for each gene in alphabetical order, and the allele functionality and metabolizer phenotype that corresponds with the genotype, if applicable. Only genes for which a genotype could be determined by PharmCAT are included in the summary table. The last column of the table indicates if variant positions evaluated by PharmCAT are missing in the provided VCF file. If positions are missing, the resulting genotype is considered to be based on incomplete information. The drugs are color/symbol coded to indicate whether the applicable CPIC guideline recommends a prescribing change based on the given genotype. The drugs are linked to the CPIC Recommendation section for quick access to drugs of interest. The genes are linked to the Allele Call Details section.
2. CPIC Recommendation. The second section lists the specific genotype-based CPIC recommendation for each drug in alphabetical order. Links to the CPIC and PharmGKB websites are provided, along with the citation of the published guideline. In the future, this section is intended to provide recommendations and annotations from other sources in addition to CPIC, and this section will be renamed accordingly at that time.
3. Allele Call Details. The third section provides a table of variant information per gene in alphabetical order. The table includes the genomic position, rsID, DNA bases reported in the VCF file, reference base and any star allele in which that variant is included. If information is not present in the VCF file for a variant evaluated by PharmCAT, the variant is highlighted as ‘missing’ in the table. For *TPMT* and *DPYD*, there are variants listed in the table which are not included

in the PharmCAT allele definitions but appear in the CPIC guideline supplements.

4. Disclaimers. The fourth section contains Disclaimers, caveats and other information from the Reporter.

Details of the discordance between PharmCAT and GeT-RM genotype (2) results

Individual genotypes by sample and gene for concordant and discordant results are found in Table S1.

CYP4F2

For *CYP4F2*, PharmCAT could not reconcile a diplotype for seven samples because the variants in the samples do not match the star allele definitions. The VCF files show that these samples appear to have the variant defining the *2 allele and the variant defining the *3 allele on the same chromosome, essentially a *3/(*2+*3) or *1/(*2+*3) diplotype. A combination of the *2 and *3 defining variants is not currently part of the existing star allele nomenclature (https://www.pharmvar.org/gene/CYP4F2, accessed 04/22/2019). The GeT-RM project reported either *3/*3 or (*2)/*3 for these samples.

The sample NA12006 is reported as *CYP4F2(*2)/*3* and the sample NA18868 as *CYP4F2*1/(*2)* by GeT-RM. PharmCAT reported the samples as *1/*3 and *1/*1, respectively. According to Table 2 of the GeT-RM publication, the *2 allele was only tested by one method, while *3 was determined by 2 or more methods (2).

DPYD

No direct concordance based on genotype names could be found for *DPYD* since GeT-RM reported star alleles for this gene while PharmCAT reports individual variants. However, the CPIC *DPYD* allele functionality table provides a mapping of the variants included in the CPIC DPYD guideline (3) to star alleles when applicable (https://www.pharmgkb.org/page/dpydRefMaterials). PharmCAT covers positions for which the DPYD guideline (3) provides recommendations that alter prescribing, including ‘no function’ variants (included in *2A, *3, *7, *8, *10, *12, *13 definitions) and ‘decreased function’ variants (not included in the star allele system) with strong or moderate evidence. ‘Normal function’ variants (included in *4, *5, *6, *9A and B, *11 definitions) are not interrogated by PharmCAT. If no ‘no function’ or ‘decreased function’ variants are detected in the VCF, PharmCAT reports prescribing guidance based on two normal function alleles. In GeT-RM, a large number of samples included ‘normal function’ alleles such as *9 and *4. *1/*9 and (*4)/*9 GeT-RM results contain no ‘no function’ or ‘decreased function’ variants, so are equivalent to a PharmGKB report of two normal function alleles. Based on this criteria, 56 of 59 samples are concordant.

The three samples that are not concordant contained ‘decreased function’ variants found by PharmCAT at positions which were not listed as interrogated positions in the GeT-RM project, therefore could not be detected by that project.

SLCO1B1

SLCO1B1 star nomenclature represents various SNPs alone and in combination. No official nomenclature source exists for this gene, instead the allele definitions used in the CPIC guideline for *SLCO1B1* are extracted from several articles (1). According to the GeT-RM tables provided in the article, the following star alleles were tested: *1A, *1B, *2, *3, *4, *5, *6, *7, *8, *9, *10, *11, *12, *13, *14, *15, *16, *17, *18, *21. However, the specific variants used to define the alleles are not listed.

PharmCAT could not reconcile a star allele diplotype for 54 of the 59 samples. Comparing the variants in the sample VCF files to the allele definitions revealed that some variants do not exclusively exist in the star allele combinations. For example, the *SLCO1B1*18* and *19 definitions include rs4149057 in combination with other variants. Rs4149057 was present in many of the samples but the other variants were not, so PharmCAT could not report *18 and *19 in these samples. Similarly, rs2291075, which is part of the *20 and *21 definitions, was detected frequently in the samples without the other variants in those allele definitions. Again, PharmCAT was unable to report diplotypes for these samples.

If PharmCAT cannot provide a match based on the *SLCO1B1* allele definitions for a sample, the genotype at position rs4149056 is reported because CPIC provides recommendations for that genotype in addition to star allele diplotypes. In order to compare GeT-RM star allele diplotypes and PharmCAT rs4149056 genotype results, the resulting phenotypes for each sample were used. The GeT-RM diplotype was converted to phenotype using CPIC's '*SLCO1B1* Diplotype-Phenotype Table'

(https://www.pharmgkb.org/page/pgxGeneRef). Comparing SLCO1B1 phenotypes between GeT-RM and PharmCAT results in concordance for 48 more samples. Importantly, samples reported by GeT-RM as *1/*15 or *1/*17 are reported by PharmCAT as heterozygous (C/T) for rs4149056, and those reported by GeT-RM as *5/*15 are reported by PharmCAT as homozygous (C/C). These results translate to the same SLCO1B1 function and CPIC recommendation.

Four samples were reported by GeT-RM as *SLCO1B1*1/*14*, which CPIC considers 'Possible Increased function' and provides no recommendation to guide simvastatin therapy (1). PharmCAT could not reconcile a diplotype for these samples, so it reported the rs4149056 T/T genotype. which CPIC considers 'Normal function' and provides a prescribing recommendation (1). Similarly, two samples were reported by GeT-RM as SLCO1B1*1/*21, which CPIC considers "Indeterminate" and provides no prescribing recommendation while PharmCAT reported rs4149056 T/T.

TPMT

For *TPMT*, one sample was reported as *1/*16 by PharmCAT and as *1/*1 by GeT-RM. Since *TPMT*16* was not part of the interrogated alleles in the GeT-RM project (2), the reported allele defaulted to *1.

UGT1A1

The CPIC guideline for atazanavir and *UGT1A1* (4) includes seven commonly tested alleles from the UDP-Glucoronosyltransferase (UGT) Alleles Nomenclature website namely *6, *27, *28, *36, *37, *60, *80. These star allele definitions are based on a single variant each, however it is known that combinations of *UGT1A1* variants on one chromosome are possible (*60 in particular has been shown to be present with a number of other *UGT1A1* variants). Star *28 and *80 are in linkage disequilibrium, although not 100%, therefore, most samples that are identified for *28 or *37 harbor the *80 variant (https://cpicpgx.org/guidelines/guideline-for-atazanavir-and-ugt1a1/). To account for this, PharmCAT lists all alleles identified in the sample and uses a specific logic to determine the allele function and metabolizer phenotype based on the alleles in the sample. Pratt et al. Table 1 shows three different genotyping assays were used to determine the GeT-RM consensus diplotypes for *UGT1A1* (2). Further alleles, not included in the CPIC definitions, can be interrogated by the selected platforms, e.g. *7. However, only *6, *27, *28, *60 were interrogated by two out of the 3 detection methods. Disregarding *80 in the PharmCAT calls, the alleles provided by GeT-RM are concordant with the alleles returned by PharmCAT in 24 cases.

Summary of concordant *UGT1A1* calls

GeT-RM - UGT1A1	PharmCAT - UGT1A1	Number of samples
*1/*1	*1/*1	4
*1/*60	*1/*60	7
*60/*60	*60/*60	2
*1/*6	*1/*6	3
*6/*6	*6/*6	1
*60 / (*28 + *60) or (*28 + *60)/*60	*60/*28+*60+*80	5
(*28 + *60) / (*28 + *60)	*28+*60+*80/*28+*60+*80	2
	total:	24

One sample was reported as *UGT1A1*1/(*7)* by GeT-RM but *UGT1A1*1/*1* by PharmCAT because *7 is not a part of the CPIC guideline or allele definition file, therefore it is not detected by PharmCAT.

For three samples, the GeT-RM results included the (*37) allele (a decreased function allele), yet the VCF files from 1000 Genomes for these samples did not include the *37 variant. Also (*36) was reported in three samples by GeT-RM but not found in the VCF files. We note that GeT-RM results in parentheses reflect that the allele was based on only one assay. Additionally, the *36, *27 and *28 alleles are defined by TA repeats of different length at the same genome position and may not be accurately reported in the NGS data/VCF file. Two samples were reported as *UGT1A1*28/*28* by GeT-RM, while the VCF files showed heterozygosity for *28, changing the UGT1A1 metabolizer phenotype from poor to intermediate for those samples.

The sample HG00436 was reported as *27/*28 by GeT-RM, however the phased VCF file showed both decreased function variants are on the same allele, again resulting in a discordance of the metabolizer phenotype. For 11 samples, GeT-RM reported *60/*28, but in the phased VCF files the variants were detected on the same allele *1/*28+*60+*80. However, in these cases the metabolizer phenotype does not change.

Summary of discordant *UGT1A1* calls

GeT-RM - UGT1A1	PharmCAT - UGT1A1	Number of samples	Reason
*1/(*7)	*1/*1	1	*7 not part of the CPIC UGT1A1 allele definitions
*60/*28 or *28/(*60)	*1/*28+*60+*80	11	Phase indicated in VCF input and variants were detected on the same allele
*27; / (*28 + *60)	*1/*27+*28+*60+*80	1	

(*28)/(*28 + *60)	*1/*28+*60+*80	1	*28 only detected heterozygous in VCFs used for PhamCAT
(*28 + *60)/(*28 + *60)	*60/*28+*60+*80	1	
(*28 + *60)/(*36 + *60) or (*36 +*60)/(*28 + *60)	*60/*28+*60+*80	2	*36 or *37 are not detected in VCFs used for PharmCAT
(*36 + *60)/*60	*60/*60	1	
(*28 + *60)/(*37 + *60)	*60/*28+*60+*80	1	
(*37)/*60	*1/*60	2	
	total:	21	

Summary of reasons for discordance by gene

Gene	Reasons for discordance
CYP4F2	• Combination of variants for allele not part of star allele nomenclature and therefore not CPIC allele definitions, CYP4F2(*2+*3). • Variant not detected in the VCFs used for PharmCAT but detected in the GeT-RM.
DPYD	• Variant detected through PharmCAT that was not interrogated in the GeT-RM.
SLCO1B1	• Not well-established star nomenclature, see text above.
TPMT	• Variant detected through PharmCAT that was not interrogated in the GeT-RM.
UGT1A1	• Allele included in Get-RM call that is not part of the CPIC allele definitions. • Phase indicated in the VCF input for PharmCAT shows variant on the same allele as compared to on 2 alleles in the GeT-RM output. • Variant only detected heterozygous in the VCFs used for PharmCAT. • Variant not detected in the VCFs used for PharmCAT but detected in the GeT-RM.

Additional Supplementary files

PharmCAT_Example _Report.pdf: An example of a full report output from PharmCAT.

Provided as a PDF file.

PharmCAT_TableS1.xlsx: The GeT-RM and PharmCAT results per sample, by gene

Provided as an Excel file

PharmCAT_TableS2.xlsx: A summary of the variants/alleles from the allele definitions that are validated with the GeT-RM samples and those that require validation
Provided as an Excel file

Supplementary References

(1) Ramsey, L.B. *et al.* The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. *Clinical pharmacology and therapeutics* **96**, 423-8 (2014).

(2) Pratt, V.M. *et al.* Characterization of 137 Genomic DNA Reference Materials for 28 Pharmacogenetic Genes: A GeT-RM Collaborative Project. *The Journal of molecular diagnostics : JMD* **18**, 109-23 (2016).

(3) Amstutz, U. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clinical pharmacology and therapeutics* **103**, 210-6 (2018).

(4) Gammal, R.S. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for UGT1A1 and Atazanavir Prescribing. *Clinical pharmacology and therapeutics* **99**, 363-9 (2016).