


RESEARCH

Open Access



Evaluating the frequency and the impact of pharmacogenetic alleles in an ancestrally diverse Biobank population

Shefali S. Verma^{1†}, Karl Keat^{2†}, Binglan Li³, Glenda Hoffecker⁴, Marjorie Risman⁵, Regeneron Genetics Center, Katrin Sangkuhl³, Michelle Whirl-Carrillo³, Scott Dudek⁵, Anurag Verma⁴, Teri E. Klein^{3,6}, Marylyn D. Ritchie^{5†} and Sony Tuteja^{4*†} 

Abstract

Background: Pharmacogenomics (PGx) aims to utilize a patient's genetic data to enable safer and more effective prescribing of medications. The Clinical Pharmacogenetics Implementation Consortium (CPIC) provides guidelines with strong evidence for 24 genes that affect 72 medications. Despite strong evidence linking PGx alleles to drug response, there is a large gap in the implementation and return of actionable pharmacogenetic findings to patients in standard clinical practice. In this study, we evaluated opportunities for genetically guided medication prescribing in a diverse health system and determined the frequencies of actionable PGx alleles in an ancestrally diverse biobank population.

Methods: A retrospective analysis of the Penn Medicine electronic health records (EHRs), which includes ~3.3 million patients between 2012 and 2020, provides a snapshot of the trends in prescriptions for drugs with genotype-based prescribing guidelines ('CPIC level A or B') in the Penn Medicine health system. The Penn Medicine BioBank (PMBB) consists of a diverse group of 43,359 participants whose EHRs are linked to genome-wide SNP array and whole exome sequencing (WES) data. We used the Pharmacogenomics Clinical Annotation Tool (PharmCAT), to annotate PGx alleles from PMBB variant call format (VCF) files and identify samples with actionable PGx alleles.

Results: We identified ~316,000 unique patients that were prescribed at least 2 drugs with CPIC Level A or B guidelines. Genetic analysis in PMBB identified that 98.9% of participants carry one or more PGx actionable alleles where treatment modification would be recommended. After linking the genetic data with prescription data from the EHR, 14.2% of participants (n = 6157) were prescribed medications that could be impacted by their genotype (as indicated by their PharmCAT report). For example, 856 participants received clopidogrel who carried *CYP2C19* reduced function alleles, placing them at increased risk for major adverse cardiovascular events. When we stratified by genetic ancestry, we found disparities in PGx allele frequencies and clinical burden. Clopidogrel users of Asian ancestry in PMBB

[†]Shefali S. Verma, Karl Keat, Marylyn D. Ritchie and Sony Tuteja: contributed equally to this work.

*Correspondence: sonyt@pennmedicine.upenn.edu

⁴ Department of Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA

Full list of author information is available at the end of the article



had significantly higher rates of *CYP2C19* actionable alleles than European ancestry users of clopidogrel ($p < 0.0001$, OR = 3.68).

Conclusions: Clinically actionable PGx alleles are highly prevalent in our health system and many patients were prescribed medications that could be affected by PGx alleles. These results illustrate the potential utility of preemptive genotyping for tailoring of medications and implementation of PGx into routine clinical care.

Keywords: Pharmacogenomics, Electronic Health Records, Genomic analyses, Genetic testing, CPIC guidelines

Introduction

The vision of precision medicine involves using genomic information to guide preventive measures and tailor treatment for patients. A major component of this vision is pharmacogenomics (PGx), which aims to provide pharmacotherapy guidance for patients carrying genetic alleles impacting response to medications, with the goal of genetically tailored prescribing at the point of care. Many medications, such as opioid analgesics or the anticoagulant warfarin, require careful dose titration to achieve a therapeutic dose, and consideration of PGx information can help select the most appropriate drug at the right dose, avoiding harmful adverse drug reactions. The Clinical Pharmacogenetics Implementation Consortium (CPIC) has published guidelines for 71 drugs based on a large body of evidence showing the impact of PGx alleles which provide specific drug recommendations to alter prescribing practices. [1–3] CPIC also assigns preliminary prioritization levels (called “provisional”) based on cursory review of evidence including, but not limited to, actionability in other professional society guidelines, recommendations from FDA-approved drug labels and assessment of evidence from PharmGKB clinical annotations. [4, 5] Level A refers to gene-drug pairs where genetic information *should be used* for prescribing decisions and alternative therapies, or dosing are highly likely to be effective and safe. [1] Level B refers to pairs where genetic information *could be used* to change prescribing and alternative therapies or dosing are extremely likely to be as effective and as safe as non-genetically based dosing. [1] As of June 2022, there are 96 gene-drug pairs with a CPIC level A or B guideline. Many of the alleles in these genes are common in the population and associated with a large effect size and a clearly identifiable drug response phenotype such as metabolizer status. However, one barrier to PGx implementation is the uncertainty about the true clinical burden of PGx alleles, particularly when accounting for ancestry-specific allele frequencies. For example the *CYP2C19**2 no function allele (with rs4244285 as a key SNP) is twice as common in Asian ancestry individuals as those with European ancestry, which will have ethical and legal implications for testing prioritization

[6] Understanding the prevalence of PGx alleles and the frequency of prescribing drugs impacted by these alleles in multiple ancestral populations will aid in targeting clinical implementation efforts for drug-gene pairs with the greatest impact.

Electronic Health Records (EHRs) are a digital record of a patient’s diagnosed conditions as well as information about the care they have received. EHRs have also become a robust source of real-world data that when linked to large DNA repositories can validate the association of known PGx alleles with drug response phenotypes but also further enable discovery of novel PGx associations. There have been several successful examples using the EHR to identify novel genetic association with drug side effects. [7–9] The University of Pennsylvania Health System (also known as Penn Medicine) is one of the primary health care providers in Philadelphia and the surrounding region. Penn Medicine consists of over 3.3 million patients whose data are stored in an EHR since 2008. In our study period of 2012–2020, we counted 3,384,922 unique patients. Approximately 150,000 of these patients have enrolled in the Penn Medicine BioBank (PMBB) research program as of July 2022 and thus far 43,359 have been genotyped for research purposes on a genome-wide SNP array and sequenced using whole exome sequencing (WES); all participant genetic data is then linked to their EHR data. Prior studies have shown that >95% of individuals carry one or more alleles that interact with drugs from CPIC Level A guidelines [10–12], but the number of patients that will actually be prescribed the drugs impacted by the alleles they carry is challenging to predict. In consideration of the implementation of preemptive PGx into clinical care, we sought to understand the population of Penn Medicine patients that are impacted by PGx alleles and are prescribed PGx affected medications. Considering this, we evaluated the prescribing patterns in the Penn Medicine EHR for a set of medications that are included in CPIC level A and B guidelines. Actionable PGx alleles are ones for which a medication prescribing change is recommended. Therefore, we also evaluated the frequency of actionable PGx phenotypes, the set of alleles that contribute to a clinically meaningful and observable drug response trait (e.g., drug metabolizer status), and are associated with strong

or moderate therapeutic recommendations in the CPIC guidelines, in the genotyped PMBB participants. The large number of ancestrally diverse participants in the PMBB also allowed for ancestry stratified analyses to better understand how drug prescription practices based on an individual's ancestry may influence the burden of unwanted drug response outcomes.

Methods

Dataset

The study was approved by the Institutional Review Board of the University of Pennsylvania and complied with the principles set out in the Declaration of Helsinki. All individuals recruited for the Penn Medicine BioBank (PMBB) are patients of clinical practice sites of the University of Pennsylvania Health System. Appropriate consent was obtained from each participant regarding storage of biological specimens, genetic sequencing/genotyping, and access to all available EHR data. Medication data was extracted using Penn Medicine's Clinical Data Warehouse, the Penn Data Store (PDS), from January 2012–December 2020. We used the CPIC guidelines available prior to December 2021 to select 49 drugs with prescribing recommendations for our study. For each of the drugs (Additional file 1: Table S1), we generated extensive medication libraries that contained every permutation of the medication available in the EHR including by drug name (generic, brand name and combinations products), dose and dosage form (e.g. oral, intravenous) and RxNorm identifiers. [13] Unique patients were counted once per year for each medication to get a unique number of prescriptions per year. Yearly prescribing data were normalized by yearly encounters to the health system to obtain a drug exposure rate and account for the growing number of patients in the health system. Drug exposures and encounters were limited to those which occurred when the patient was 18 years of age or older.

Genotyping and whole exome sequencing

Individuals were genotyped using the Illumina Global Screening array v.2.0 and minimally processed with PLINK to remove sites with marker call rate <95% and samples with overall call rate <90%. Furthermore, any samples with a discordance between genetically defined sex and EHR reported sex were discarded. Genotyping and whole exome sequencing was performed at the Regeneron Genetics Center using an in-house high throughput, fully automated approach for exome capture. Libraries were generated from 100 ng of genomic DNA with a mean fragment length of 200 bp. The libraries were then barcoded and amplified in preparation for

multiplexed exome capture and sequencing. Following exome capture, pooled samples were sequenced with 75 bp paired end reads with two 10 bp index reads on the Illumina NovaSeq 6000 platform on S4 flow cells. Samples were then QC filtered using the widely used 'Goldilocks' filtering procedure as described previously [14, 15].

Construction of an integrated call set

Genotyping arrays and exome sequencing in isolation both provide limited coverage of PGx sites across the genome. We took a similar, but simplified approach to McInnes et al. to generate an integrated call set combining the array and exome data. [16] Based on prior work indicating that whole exome sequencing has a much higher call accuracy than genotyping arrays [17], the integrated call set favored exome calls over array calls in cases where a site was covered by both. The exome was used as the primary sequence, with sites not covered by the exome filled with available calls from the genotyping array using bcftools version 1.15.1 [18]. Average coverage was assessed for the array, exome, and integrated call sets by counting the fraction of PharmCAT reference PGx sites that were found in each set.

PGx analysis

Annotation of PGx alleles was performed using an in-development version of PharmCAT 2.0, a bioinformatics tool to annotate PGx alleles from a VCF file [19]. The version of PharmCAT used for this study interprets alleles included in CPIC guideline genes, using the latest CPIC guidelines published in August 2022. The PharmCAT preprocessor was used to convert the PMBB multi-sample VCF files into single-sample VCF files with normalized allele representations for input into the main PharmCAT tool. PharmCAT-generated annotations for the 43,359 genotyped individuals were used to identify all PMBB individuals with one or more actionable PGx alleles for the studied drugs.

PharmCAT provides a report for each individual (patient-participant) containing all drug-gene combinations that have guidelines included in the CPIC database with a few exceptions. The version of PharmCAT at the time of this study did not call *G6PD*, which falls on the X-chromosome, as well as *MT-RNR1*, a mitochondrial gene and also did not call the *HLA* region. While we studied the usage rates of *some* drugs which are affected by *HLA* and *G6PD* genotype/phenotype, we could not study the frequencies of PGx alleles in these genes. CPIC guidelines utilize phenotypes to provide dosing recommendations rather than genotypes. As such, PharmCAT performs genotypes to phenotype conversions based on allele function guidelines from CPIC. These phenotype frequencies

are provided in Table 3. Phenotypes are usually listed as poor metabolizer (PM), intermediate metabolizer (IM), normal metabolizer (NM), rapid metabolizers (RM), and ultrarapid metabolizer (UM). For some complex genes such as *CYP2D6*, genotypes are translated into gene activity scores (AS) to better capture residual enzyme activity for some alleles [3]. We calculated the AS for *CYP2C9* and *CYP2D6* based on the CPIC allele functionality tables. Using the PharmCAT calls, we performed an ancestry-stratified analysis to identify patterns of allele and phenotype frequency in different populations. Furthermore, for each drug-gene combination we looked at the rates of patients with actionable PGx alleles (i.e., leading to prescribing change recommendations) for drugs which they were prescribed, stratified by ancestry. A phenotype is deemed actionable if there is a CPIC guideline for that phenotype and a particular drug. In some cases, such as *CYP2C19*, which has both gain-of-function and loss-of-function variants in the CPIC definitions, some non-baseline phenotypes may be actionable for one drug but not another. For instance, *CYP2C19* UM phenotypes have a CPIC guideline recommendation for omeprazole, but not for clopidogrel. Therefore, while *CYP2C19* UM is an actionable *CYP2C19* phenotype, it is not actionable for clopidogrel specifically, which instead has guidelines for *CYP2C19* IM and PM.

PharmCAT does support inclusion of external calls for *HLA*, *G6PD*, and *CYP2D6*. Due to the limitations of array and exome data, we were unable to call *CYP2D6* CNV and structural variation with any external tool. PharmCAT does not ordinarily call the *CYP2D6* gene, whose phenotype is reliant on copy-number and structural variation in the gene as well as the *HLA* locus which is difficult to call without specialized assays. However, a “research mode” option in PharmCAT allowed us to call *CYP2D6* from SNPs alone. These calls only feature *CYP2D6* star-alleles that do not define or include structural variation or a gene deletion. We conducted a separate analysis using the *CYP2D6* calls to estimate the prevalence of *CYP2D6* reduced function phenotypes in patients treated with

relevant drugs, with the notable caveat that we could not measure effects of gene deletion or duplication.

Results

Demographics

There were 11,235,263 unique patient encounters to Penn Medicine from January 1, 2012, to December 31, 2020, across 1,896,012 unique patients. The cohort included 56.1% women and 43.9% men. The median age was 56 years. The EHR race and ethnicity demographics were 59.8% White (56.7% self-reported White and 3.2% self-reported Hispanic White), 23.4% Black (22.4% self-reported Black and 1% self-reported as Hispanic Black), 4.2% Hispanic or Latino, and 4% Asian as shown in Additional file 1: Fig. S1a.

The distribution of genetic ancestry in the PMBB is comprised of 69.2% European ancestry (EUR), 25.7% African ancestry (AFR), 1.6% East Asian ancestry (EAS), 1.3% admixed American ancestry (AMR), 1.3% South Asian ancestry (SAS), and 0.9% unknown ancestry (UNK). 49.7% of PMBB patients are female (Additional file 1: Fig. S1b). Differences in the demographics of PMBB and Penn Medicine can be partly attributed to the distinction between genetic ancestry and self-reported race, as well as the recruitment of most genotyped PMBB patients occurring prior to Penn Medicine’s acquisition of multiple hospitals.

Drug prescription trends in Penn Medicine EHR

During our study period, 723,115 (21.3%) unique patients receiving care at Penn Medicine had at least 1 record for a medication in our list (see “Methods” section), while 316,235 (9.3%) received 2 or more drugs and 146,885 (4.3%) received 3 or more. The count of medications per patient per year is show in Table 1. The most prescribed medication over the study period was ondansetron, with more than 3 ondansetron prescriptions for every 100 encounters. Ondansetron is an anti-emetic medication used to treat nausea and vomiting (Table 2).

An evaluation of prescribing patterns over time for the studied medications showed an increasing use of some medications, particularly for ondansetron and ibuprofen

Table 1 Total number of medications taken by unique patients each year from 2012–2020

# Meds	2012	2013	2014	2015	2016	2017	2018	2019	2020
0	620.820	605.509	593.552	568.411	550.173	525.943	504.553	489.512	508.213
1	74.261	84.061	92.561	114.240	126.647	139.960	156.680	167.905	155.969
2	20.338	24.040	26.197	29.654	33.772	40.074	44.121	47.179	42.731
3	5774	6954	7885	7985	9188	11.963	12.686	13.219	11.720
4	1481	1923	2177	2135	2513	3687	3653	3800	3300
5+	441	628	743	690	822	1488	1422	1500	1182

Table 2 List of PGx informed medications and total prescriptions for each medication. Prescription rate is calculated as total prescriptions/number of unique patient encounters. The list is ordered from high to low prescription rate

Medication	Drug class	Total prescriptions	Prescription rate
Ondansetron	Anti-emetic	387.151	0.034459
Ibuprofen	NSAID	382.539	0.034048
Omeprazole	Proton-pump inhibitor	364.753	0.032465
Simvastatin	Statin	224.908	0.020018
Sertraline	Antidepressant	187.200	0.016662
Tramadol	Opioid	184.633	0.016433
Escitalopram	Antidepressant	170.216	0.01515
Codeine	Opioid	152.598	0.013582
Pantoprazole	Proton-pump inhibitor	146.450	0.013035
Meloxicam	NSAID	93.665	0.008337
Citalopram	Antidepressant	92.664	0.008248
Clopidogrel	Antiplatelet	86.931	0.007737
Warfarin	Anticoagulant	84.435	0.007515
Hydrocodone	Opioid	82.405	0.007334
Lansoprazole	Proton-pump inhibitor	79.480	0.007074
Tacrolimus	Immunosuppressant	67.348	0.005994
Allopurinol	Uric acid reducer	62.078	0.005525
Celecoxib	NSAID	51.544	0.004588
Amitriptyline	Antidepressant	41.721	0.003713
Paroxetine	Antidepressant	40.524	0.003607
Nortriptyline	Antidepressant	29.154	0.002595
Azathioprine	Immunosuppressant	26.068	0.00232
Fluorouracil	Chemotherapy	25.805	0.002297
Tamoxifen	Chemotherapy	15.005	0.001336
Oxcarbazepine	Anticonvulsant	14.975	0.001333
Carbamazepine	Anticonvulsant	14.116	0.001256
Dexlansoprazole	Proton-pump inhibitor	10.077	0.000897
Abacavir	Antiviral	9709	0.000864
Doxepin	Antidepressant	9175	0.000817
Efavirenz	Antiviral	8856	0.000788
Phenytoin	Anticonvulsant	7475	0.000665
Piroxicam	NSAID	5665	0.000504
Atomoxetine	ADHD	5003	0.000445
Flurbiprofen	NSAID	4802	0.000427
Capecitabine	Chemotherapy	4672	0.000416
Atazanavir	Antiviral	4099	0.000365
Voriconazole	Antifungal	4037	0.000359
Mercaptopurine	Chemotherapy	3601	0.000321
Fluvoxamine	Antidepressant	3440	0.000306
Desipramine	Antidepressant	2458	0.000219
Imipramine	Antidepressant	2050	0.000182
Clomipramine	Antidepressant	1741	0.000155
Succinylcholine	Paralytic	1373	0.000122
Ivacaftor	Cystic Fibrosis	1293	0.000115
Peginterferon_alfa-2a	Antiviral	993	8.84E-05
Rasburicase	Uric acid reducer	520	4.63E-05
Peginterferon_alfa-2b	Antiviral	322	2.87E-05
Fosphenytoin	Anticonvulsant	109	9.70E-06

Table 2 (continued)

Medication	Drug class	Total prescriptions	Prescription rate
Thioguanine	Chemotherapy	20	1.78E-06

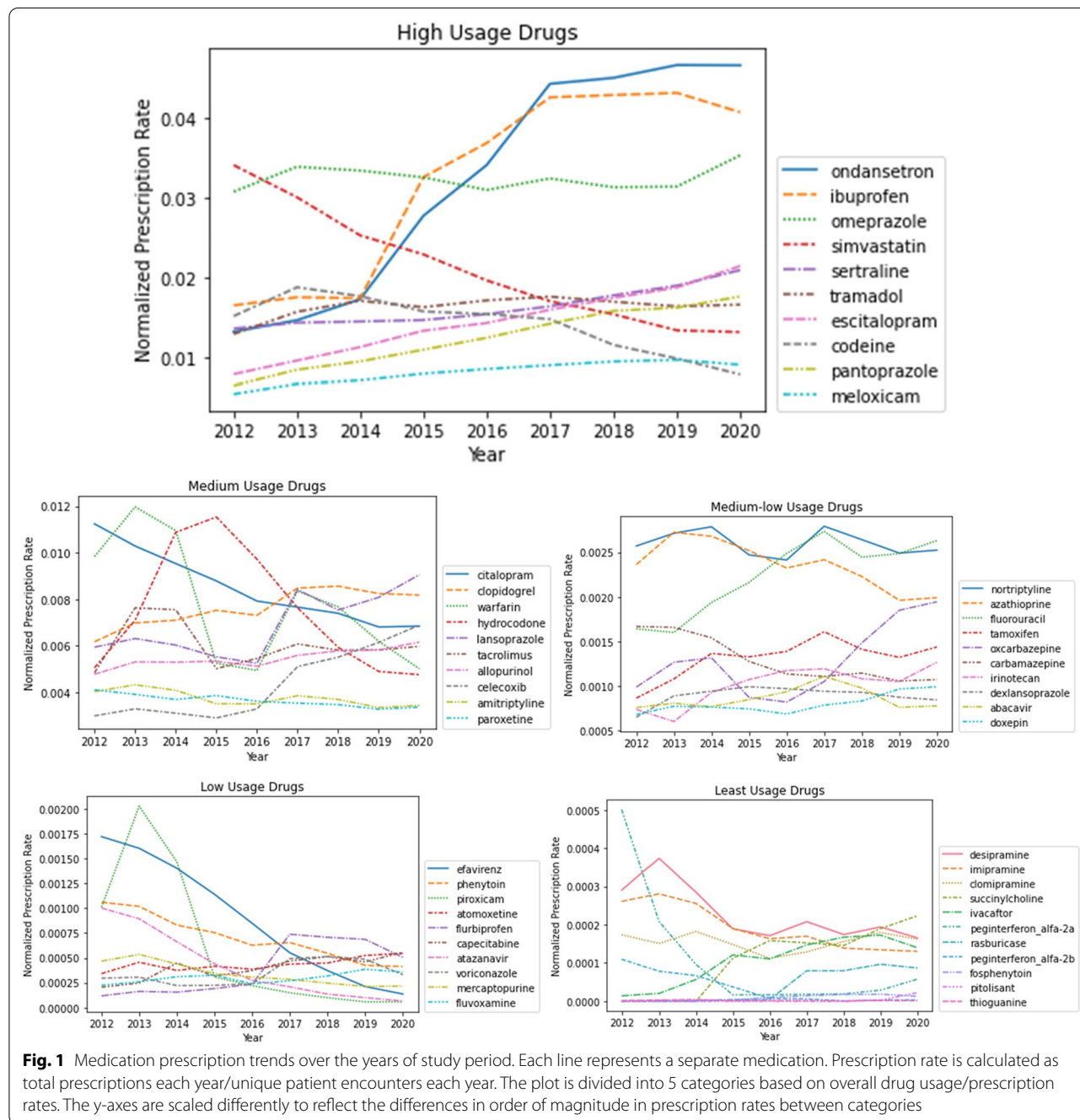
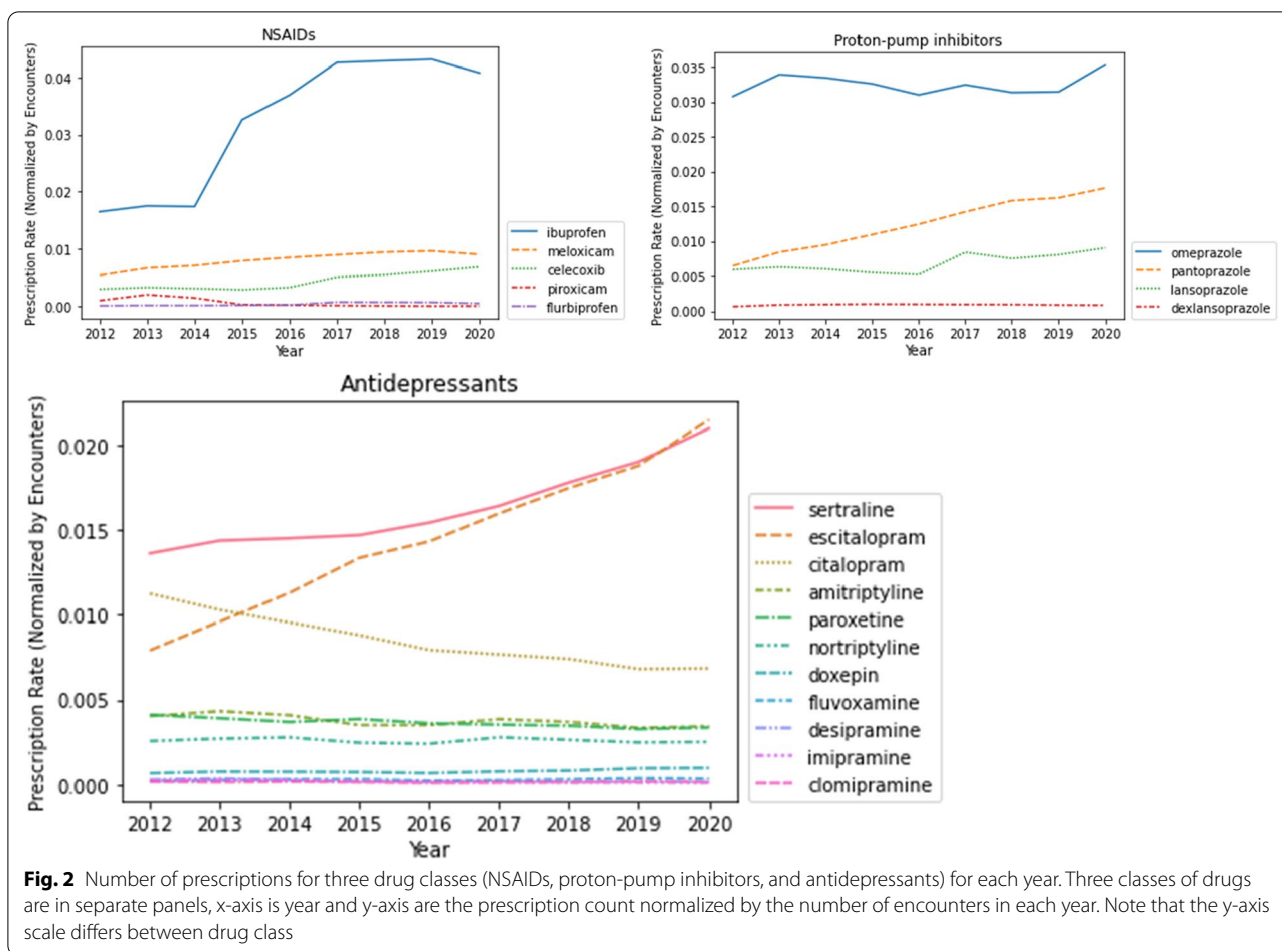


Fig. 1 Medication prescription trends over the years of study period. Each line represents a separate medication. Prescription rate is calculated as total prescriptions each year/unique patient encounters each year. The plot is divided into 5 categories based on overall drug usage/prescription rates. The y-axes are scaled differently to reflect the differences in order of magnitude in prescription rates between categories

when the total prescriptions are normalized by the number of encounters for each year as shown in Fig. 1. At the same time, despite a large decline in the usage of

simvastatin, it remains one of the most prescribed drugs, with more than 1 in 100 encounters involving a prescription. Similarly, while the use of warfarin is declining, it is



still used very frequently. Additionally, we examined prescribing patterns for 3 drug classes: non-steroidal anti-inflammatory drugs (NSAIDs), proton-pump inhibitors, and antidepressant drugs (Fig. 2).

Integrated call set performance

We compared the performance of our integrated call set to the whole exome sequencing and genotyping array data by measuring their coverage of the PGx sites used by PharmCAT. Of 536 total sites across all samples, the arrays covered an average of 250 sites (46.6%), the exome covered an average of 259 (48.3%) and the integrated call set covered an average of 354 (66%). This validates that exome and genotyping arrays have complementary information for PGx and should be combined, when possible, for larger coverage of the genome. For example, the noncoding variant *rs12248560* is vital to calling the *CYP2C19**17 increased function allele, which increases *CYP2C19* gene expression. In combination with a normal function allele, the *17 allele results in a RM phenotype. In case of two *17 alleles, there is an UM phenotype.

Lastly, combinations of *17 with reduced function or loss of function alleles can result in IM phenotypes. Since *rs12248560* was not included in the exome capture, it had to be filled in using the array data.

Gene-drug interaction analyses

We evaluated number of gene-drug interactions within the PMBB with medications with CPIC level A or B guidelines available. We extracted 16 loci (15 genes + the *rs12777823* SNP) as listed in Table 3 that have gene-drug interactions. Among the ~43,000 genotyped individuals, all individuals (100%) have at least one non-reference allele in the subset of loci we genotyped. PharmCAT annotation also identified that 98.9% of individuals are carriers of one or more PGx actionable phenotypes that would result in a pharmacotherapy modification. The mean number of genes per patient with an actionable phenotype was 4.1. A t-test comparing EUR ($\mu=4.088$, $n=30,008$) and AFR ($\mu=4.282$, $n=11,156$) actionable phenotype counts showed that AFR has a significantly higher mean count of phenotypes that correspond to treatment modification than EUR ($p<0.0001$). The

Table 3 Number of participants in the PMBB with PGx phenotypes by ancestry based on PharmCAT annotation

Gene	Phenotype	Genetic Ancestry						TOTAL
		EUR	AFR	EAS	SAS	AMR	UNK	
CACNA1S	Malignant Hyperthermia Susceptibility	4	0	0	0	0	0	4
	Uncertain Susceptibility	30.000	11.154	676	562	580	376	43.348
	No Result	4	2	1	0	0	0	7
CFTR	Ivacaftor responsive in CF patients	474	337	3	7	16	4	841
	Ivacaftor non-responsive in CF patients	29.512	10.808	674	555	564	372	42.485
	No Result	22	11	0	0	0	0	33
CYP2B6	Intermediate Metabolizer	10.475	5204	222	253	241	142	16.537
	Ultrarapid Metabolizer	30	11	3	3	0	0	47
	Ambiguous	9	0	0	0	0	0	9
	Rapid Metabolizer	965	266	57	19	15	17	1339
	Indeterminate	1398	149	2	3	13	7	1572
	Normal Metabolizer	15.056	3086	350	199	235	155	19.081
	Poor Metabolizer	1776	2060	34	80	69	48	4067
	No Result	299	380	9	5	7	7	707
CYP2C19	Intermediate Metabolizer	7538	3465	322	246	119	119	11.809
	Ultrarapid Metabolizer	1389	542	0	29	12	12	1984
	Likely Poor Metabolizer	3	71	0	0	1	1	76
	Likely Intermediate Metabolizer	25	276	0	0	6	2	309
	Rapid Metabolizer	8092	2603	9	90	126	68	10.988
	Indeterminate	7	278	0	1	3	2	291
	Normal Metabolizer	12.208	3441	245	122	305	153	16.474
	Poor Metabolizer	692	423	98	67	8	19	1307
CYP2C9	No Result	54	57	3	7	0	0	121
	Intermediate Metabolizer	6889	2369	11	48	110	71	9498
	Ambiguous	4125	341	38	107	56	43	4710
	Indeterminate	769	204	17	21	19	13	1043
	Normal Metabolizer	18.170	8177	611	386	395	247	27.986
	Poor Metabolizer	7	33	0	0	0	0	40
	No Result	48	32	0	0	0	2	82
CYP3A5	Intermediate Metabolizer	3898	5512	294	208	167	110	10.189
	Poor Metabolizer	25.959	3283	333	309	388	247	30.519
	Indeterminate	2	1	0	0	0	0	3
	Normal Metabolizer	146	2357	50	45	25	19	2642
	No Result	3	3	0	0	0	0	6
CYP4F2	Not actionable	16.837	9289	420	257	339	221	27.363
	Actionable	13.171	1867	257	305	241	155	15.996
DPYD	Intermediate Metabolizer	2028	676	16	50	21	18	2809
	Ambiguous	8	7	0	0	1	0	16
	Indeterminate	642	511	6	23	16	10	1208
	Normal Metabolizer	27.283	9952	655	487	542	348	39.267
	Poor Metabolizer	47	10	0	2	0	0	59
IFNL3/4	not actionable	13.637	2078	571	345	245	196	17.072
	actionable	16.371	9078	106	217	335	180	26.287

Table 3 (continued)

Gene	Phenotype	Genetic Ancestry						TOTAL
		EUR	AFR	EAS	SAS	AMR	UNK	
NUDT15	Intermediate Metabolizer	317	36	103	53	40	20	569
	Indeterminate	163	77	38	1	13	7	299
	Normal Metabolizer	29,526	11,043	524	506	527	349	42,475
	Poor Metabolizer	1	0	6	2	0	0	9
	Possible Intermediate Metabolizer	0	0	6	0	0	0	6
	No Result	1	0	0	0	0	0	1
RYR1	Malignant Hyperthermia Susceptibility	23	11	0	0	0	0	34
	Uncertain Susceptibility	29,979	11,143	676	562	580	376	43,316
	No Result	6	2	1	0	0	0	9
TPMT	Intermediate Metabolizer	2293	1068	17	18	68	36	3500
	Indeterminate	466	995	16	10	15	11	1513
	Normal Metabolizer	27,118	9007	643	533	493	325	38,119
	Poor Metabolizer	54	34	0	0	1	1	90
	Possible Intermediate Metabolizer	61	48	1	1	1	3	115
	No Result	16	4	0	0	2	0	22
UGT1A1	Intermediate Metabolizer	14,594	5708	274	316	276	182	21,350
	Ambiguous	11	6	0	0	0	0	17
	Indeterminate	1322	702	3	44	24	6	2101
	Normal Metabolizer	11,966	3233	328	140	232	146	16,045
	Poor Metabolizer	1528	1199	67	44	37	31	2906
	No Result	587	308	5	18	11	11	940
VKORC1	Actionable	19,264	2327	663	183	389	252	23,078
	Not actionable	10,744	8829	14	379	191	124	20,281
SLCO1B1	Increased Function	1280	155	0	1	9	5	1450
	Poor Function	825	66	11	5	7	10	924
	Possible Decreased Function	28	160	0	0	0	4	192
	Ambiguous	479	29	0	0	6	1	515
	Normal Function	19,382	7234	509	452	408	243	28,228
	Indeterminate	250	2279	14	39	7	12	2601
	Decreased Function	7764	1233	143	65	143	101	9449
rs12777823	Not actionable	29,406	10,501	618	499	572	359	41,955
	Actionable	602	655	59	63	8	17	1404

*Indeterminate means the called genotype did not have an assigned function in the CPIC guidelines. No Result means that the genotype could not be called due to missing or invalid positions

number of individuals stratified by each ancestry group for gene-phenotype combination are listed in Table 3. Next, we examined how many individuals with actionable alleles were prescribed drugs that would be impacted by those alleles as represented in Table 4. Among the top prescribed CPIC drugs where prescribed individuals have an actionable PGx allele are warfarin, clopidogrel, and omeprazole.

Since our data show a higher frequency of *CYP2C19* reduced function alleles in East Asian ancestry individuals, we decided to see if this trend resulted in larger

proportions of clopidogrel users of East Asian ancestry having actionable *CYP2C19* IM and PM phenotypes. We performed a Fisher's exact test to compare these proportions of individuals of EUR and Asian ancestry (EAS & SAS) prescribed clopidogrel who had IM or PM phenotypes. In the EUR group there were 537 IMs and PMs (actionable for clopidogrel), and 1451 patients with nonactionable phenotypes. In the Asian ancestry group, there were 22 without and 30 with actionable phenotypes. We found a significantly larger proportion of Asian ancestry individuals taking clopidogrel are IMs and PMs

Table 4 Top 10 drugs by number of PMBB participants with PGx actionable alleles. Proportion of participants prescribed a medication also carrying an actionable allele for that drug. The count of prescribed drugs with actionable alleles (ACT), total prescribed drugs considered (TOT), and proportion of prescribed drugs with actionable alleles (PROP) are shown

Drug	Genes	Proportion of Total Prescribed with Actionable PGx																				
		EUR		AFR		EAS		SAS		AMR		UNKNOWN		TOTAL								
		ACT	TOT	PROP	ACT	TOT	PROP	ACT	TOT	PROP	ACT	TOT	PROP	ACT	TOT	PROP						
<i>warfarin</i>	CYP2C9 CYP4F2 VKORC1 rs12777823	1904	2216	0.8592	416	864	0.4815	19	20	0.9500	15	18	0.8333	19	22	0.8636	23	26	0.8846	2396	3166	0.7568
<i>clopidogrel</i>	CYP2C19	537	1988	0.2701	276	753	0.3665	10	18	0.5556	20	34	0.5882	3	25	0.1200	10	19	0.5263	856	2837	0.3017
<i>tacrolimus</i>	CYP3A5	214	1494	0.1432	409	585	0.6991	15	23	0.6522	7	27	0.2593	17	39	0.4359	1	24	0.0417	663	2192	0.3025
<i>simvastatin</i>	SLCO1B1	507	1806	0.2807	94	994	0.0946	11	28	0.3929	3	24	0.1250	3	27	0.1111	7	23	0.3043	625	2902	0.2154
<i>amitriptyline</i> (CYP2D6 not used)	CYP2C19	219	511	0.4286	196	522	0.3755	0	7	0.0000	3	4	0.7500	2	15	0.1333	4	9	0.4444	424	1068	0.3970
<i>omeprazole</i>	CYP2C19	237	4780	0.0496	149	3177	0.0469	0	79	0.0000	6	105	0.0571	3	121	0.0248	2	63	0.0317	397	8325	0.0477
<i>ibuprofen</i>	CYP2C9	76	4178	0.0182	110	4137	0.0266	0	150	0.0000	0	116	0.0000	1	148	0.0068	3	113	0.0265	190	8842	0.0215
<i>escitalopram</i>	CYP2C19	115	1619	0.0710	53	648	0.0818	4	20	0.2000	3	16	0.1875	1	33	0.0303	1	12	0.0833	177	2348	0.0754
<i>voriconazole</i>	CYP2C19	140	247	0.5668	24	34	0.7059	1	1	1.0000	3	5	0.6000	3	5	0.6000	0	3	0.0000	171	295	0.5797
<i>lansoprazole</i>	CYP2C19	93	1890	0.0492	58	1086	0.0534	0	26	0.0000	0	31	0.0000	1	34	0.0294	1	28	0.0357	153	3095	0.0494

than in EUR ($p < 0.0001$) with an odds ratio of 3.68, indicating that clopidogrel users of Asian ancestry are 3.68 times more likely than individuals of European ancestry to have actionable PGx alleles for clopidogrel.

Not all genes could be genotyped unambiguously for all patients due to missing alleles. PharmCAT provides output as several genotypes with the highest matching score. In many cases, the ambiguous genotypes (i.e. two or more genotypes) result in the same phenotype. In cases when they do not, the phenotype is ambiguous as well. For example, 10.9% of patients had an ambiguous *CYP2C9* phenotype, which likely resulted in an underestimate of patients with actionable *CYP2C9* phenotypes taking ibuprofen, the second most common PGx drug in Penn Medicine. A full breakdown of uncalled or ambiguous genotypes and phenotypes from PharmCAT is available in Additional file 1: Table S2.

CYP2D6 analysis

Despite a lack of copy-number and structural variant information for *CYP2D6*, SNP-based genotyping may still provide some estimates into the burden of *CYP2D6* IM and PMs at the population level with the caveat that a number of no function alleles (defined by structural variants and the gene deletion *5) cannot be captured and samples including these are determined as IM or NM depending on the second allele. PharmCAT was able to call *CYP2D6* normal and reduced function phenotypes (NM/IM/PM) for 14,581 (33.6%) of the PMBB individuals; all other individuals were either Indeterminate due to the presence uncertain or unknown function alleles, or No Result for un-callable individuals (Additional file 1: Table S3). Of those called, 7,254 (16.7%) had reduced function phenotypes (IM/PM). If we calculate gene activity scores from diplotypes, only 14,460 (33.3%) are unambiguous since metabolizer phenotypes encompass a range of activity scores and some individuals with multiple possible genotypes can have multiple activity scores which all fall under the same metabolizer category (Additional

file 1: Table S4). Using these phenotypes, we provided a lower-bound estimate of the number of patients receiving drugs for which they have an actionable *CYP2D6* IM/PM phenotype (Table 5). However, in the absence of copy number variation, we could not call RM or UM phenotypes, which are also actionable for many *CYP2D6*-influenced drugs. Despite these limitations, we were able to detect 792 patients prescribed codeine and 761 patients prescribed tramadol with an actionable IM/PM phenotype in *CYP2D6*. For ondansetron, our most prescribed PGx-affected drug, only the UM phenotype is considered actionable and therefore we could not assess this drug-gene interaction.

Discussion

The identification of many clinically actionable alleles and their interactions with medications has led investigators to critically think about implementing PGx testing of these alleles into clinical practice. Our study examined approximately 3.3 million individuals from Penn Medicine and identified that 21.3% of patients have been prescribed at least one CPIC guideline medication and 9.3% were prescribed two or more. This evidence suggests that the clinical testing of pharmacogenetic alleles could help in informing patient care at Penn Medicine and other large healthcare systems. Our analyses on ~43,000 genotyped individuals from PMBB demonstrate that the frequency of participants with one or more clinically actionable alleles is very high. 100% of participants in our biobank population are carriers of at least one non-referent allele in a gene with a CPIC Level A or B guideline, out of which 98.9% of participants have a PGx phenotype that would lead to change in the prescription or dosage of at least one drug. More than 14% of patients were prescribed a drug for which they carry an actionable allele in our 8-year study period (2012–2020). These results suggest strong evidence for the benefits of integrating pharmacogenomic testing with current clinical practice for precision health care. Pharmacogenomic testing can be beneficial in reducing the number of drug adverse effects and nonresponse outcomes as well as minimize trial and error for dose specification. As more pharmacogenetic associations are discovered for more drugs, and as existing PGx-affected medications continue to grow in usage, the utility and clinical impact of pharmacogenomic infrastructure will only grow. At scale, these interventions can be a powerful and cost-effective method for improving patient care.

In this study, we aimed to observe the landscape of prescription trends of drugs that have a pharmacological impact in the Penn Medicine EHR to make necessary decisions about where and how to implement pharmacogenomics. Similar findings have also been reported

Table 5 Proportion of PMBB participants with actionable *CYP2D6* IM/PM phenotypes from PharmCAT research-mode annotation for drugs they have been prescribed

Drug	Actionable	Total prescribed	Proportion
amitriptyline	160	1068	0.1498
atomoxetine	29	77	0.3766
codeine	792	5278	0.1501
nortriptyline	105	865	0.1214
paroxetine	9	489	0.0184
tamoxifen	42	304	0.1382
tramadol	761	5194	0.1465

by several other previous studies [11, 12, 16, 20] but to our knowledge, our study is unique in analyzing pharmacogenomic alleles among already genotyped individuals across multiple ancestry groups for identifying individuals that could benefit from pharmacotherapy. We evaluated the potential impact of PGx testing by retrospective analyses in our PMBB population and compared our findings to similar studies in one or multiple health systems. We identified different challenges in collecting and mining data from the EHR to observe these trends. Our analyses aimed at identifying individuals who could benefit from pharmacogenomic testing while highlighting essential considerations for clinicians and researchers for implementing pharmacogenomics in clinical practice. Genotyping individuals with a one-time panel of PGx genes would impact medication prescribing over the course of care. Among the most prescribed medication classes affected by PGx alleles include pain medications, anti-emetics, and proton pump inhibitors. These medications are often prescribed in the post-operative or acute care setting and represent a high-risk patient population that will benefit from implementation of pre-emptive PGx panel testing [21]. In one retrospective analysis in patients who underwent panel based PGx testing, the presence of a gene-drug interaction for CPIC guideline medications was associated with an increased risk of 90-day hospital readmission by more than 40% [22]. An ongoing prospective study will determine whether PGx panel testing is feasible in the acute care setting [23].

Our study examined the evidence based CPIC guidelines and applied them in a clinical setting to quantify their burden on actual patients. Based on the widespread prevalence of PGx alleles and PGx-affected medication, there would be substantial benefit to preemptive, universal PGx testing in both patient outcomes and healthcare costs. Warfarin is a prime example for a drug in widespread use where genetic information may inform dosing in a large proportion of patients. Prior knowledge of warfarin responsiveness can reduce trial and error in dose titration towards a safe and effective dose. Three randomized controlled trials of PGx guided warfarin dosing have been performed and one found PGx testing to increase the time in therapeutic range [24] and one of the studies showed an improvement in the composite clinical outcome of major bleeding, supratherapeutic International Normalized Ratio (INR), venous thromboembolism, or death [25]. While the use of reactive PGx testing may not be clinical feasible, the use of preemptive testing for PGx genes to guide cardiovascular medications including warfarin was found to be cost-effective [26]. In addition to dose-modifying effects, pharmacogenomic analyses have identified many gene-drug

interactions such as clopidogrel where patient carrying an alternate allele is entirely indicated against taking a drug due to high risk of nonresponse leading to adverse outcomes. Nonresponse to clopidogrel treatment resulting in stroke or myocardial infarction is a severe outcome for patients that leads to dramatic economic costs but is preventable. Information on an individual patient's drug metabolizer status can aid clinicians in making informed decisions regarding which drug regimen should be selected for patients based on their genetic information. Owing to the ancestry diversity of the PMBB, we were able to demonstrate that the burden of PGx alleles disproportionately affects certain ancestry groups. We report notable differences in phenotype frequency for many of the studied genes in each ancestry group. In aggregate, European ancestry individuals have significantly lower average counts of actionable alleles than individuals of African ancestry ($p < 0.0001$), as well as compared to all non-European ancestry combined. Furthermore, patients of Asian ancestry taking clopidogrel therapy have greatly increased rates of treatment modifying *CYP2C19* reduced metabolizer phenotypes than individuals of European ancestry, quantifying a burden of PGx in actual patients. IM and PM phenotypes may result in increased risk of major adverse cardiovascular events in clopidogrel users, and therefore this demonstrates an unequal burden of PGx-related harms in clopidogrel patients of Asian ancestry. We found similar trends of differential PGx burden across ancestry groups for other widely used drugs including tacrolimus and voriconazole. Despite the observation of ancestry-specific patterns in PGx, we strongly discourage the use of ancestry as a proxy for PGx testing. In the case of clopidogrel, although a majority of clopidogrel-prescribed patients of Asian ancestry have actionable PGx alleles in *CYP2C19*, 42.3% still do not. Furthermore, in clopidogrel-prescribed patients of European ancestry, a still-notable 27.0% of patients have actionable *CYP2C19* alleles. These findings demonstrate how inaction towards the implementation of pharmacogenetic testing harms all patients, but disproportionately harms patients of non-European ancestry.

Although our study highlighted a large proportion of pharmacogenomic alleles with CPIC guidance in the PMBB population, several limitations should be noted. We identified variations based on an integrated call set of genotype chip and whole exome data in VCF format, which does not capture PGx-relevant copy number or structural variations (in *CYP2D6*) and is also not well suited to calling *HLA* polymorphisms. As a result, we used presumptive *CYP2D6* calls from PharmCAT which must be interpreted carefully. Another limitation is that although we were able to collect data on the prescribed

medications during encounters at Penn Medicine, we are unable to confirm if the prescription is new or was prescribed at another clinical care setting. Thus, identification of patients with adverse reactions to drugs with CPIC Level A or B guidelines from the EHR has also been challenging. To address this, future work could involve developing algorithms for identifying previously administered drugs by mining through patient notes and EHR fields, and mining through longitudinal EHR data, allergy fields, and notes for adverse reactions and nonresponse to drugs. This would allow us to further quantify the burden of PGx as a measure of actual health outcomes. It should also be noted that although the integrated call set allowed us to genotype and average of 66% of PGx sites included in CPIC Level A or B guidelines, the missing sites in some cases limit our ability to call phenotypes resulted in some patients having missing or ambiguous calls for certain genes. Targeted PGx panels or whole-genome sequencing could largely avoid this issue, but targeted panels would eventually become obsolete as new alleles are discovered. Our study is unique in that it specifically explores how ancestry-specific patterns in PGx allele frequencies are reflected in patients receiving drugs for which they have a gene-drug interaction. Although we have more than 500 patients in every studied ancestry group which is enough to confidently demonstrate differences in phenotype frequencies, the relatively small proportion of patients who have been prescribed each drug limits our power to draw conclusions about ancestry-specific drug-gene interactions. As more patients are enrolled and genotyped by PMBB, and other data becomes available in resources like All of Us, we will gain more power to disentangle these ancestry-specific associations, which carry major implications for the effect of PGx across different ancestry groups [27]. It is important to consider relevant PGx alleles in all ancestry groups as more health systems consider clinical implementation of PGx; broad, diverse inclusion of multiple ancestry groups in clinical implementation has the potential to reduce health disparities. Lastly, the CPIC guidelines are easily accessible, but implementation of these guidelines poses several challenges in a clinical setting. One important challenge is the ability to return timely genotyping results prior to prescription of the drug. Based on our findings, it seems appropriate to perform genetic testing on patients preemptively and universally before they are prescribed a relevant PGx-informed drug. However, this poses many challenges which include but are not limited to cost effectiveness, insurance coverage, future discovery of new actionable genes and alleles, integration of real time clinical decision support, availability of resources, patient-prescriber education, and review of genetic reports among others.

Conclusions

With the increasing availability of clinical decision support tools like PharmCAT, it will soon be possible to add pharmacogenetic information to EHR systems such as Epic. With proper education and resources, along with increased adoption of PGx testing, healthcare professionals will be able to utilize this information to make genetics-informed drug dosing and prescription. Addressing these barriers will be imperative before widespread adoption of clinical PGx can be achieved. Implementation of PGx into clinical workflows has the potential to affect clinical care for a very large proportion of the health care population and is a significant contributor to fully realizing the practice of precision medicine.

Abbreviations

PGx: Pharmacogenomics; CPIC: The Clinical Pharmacogenetics Implementation Consortium; EHRs: Electronic Health Records; PMBB: The Penn Medicine BioBank; SNP: Single Nucleotide Polymorphism; WES: Whole Exome Sequencing; PharmCAT: Pharmacogenomics Clinical Annotation Tool; VCF: Variant call format; PDS: Penn Data Store; PM: Poor Metabolizer; IM: Intermediate Metabolizer; NM: Normal Metabolizer; RM: Rapid Metabolizers; UM: Ultrarapid Metabolizer; EUR: European ancestry; AFR: African ancestry; EAS: East Asian ancestry; AMR: Admixed American ancestry; SAS: South Asian ancestry; UNK: Unknown ancestry; NSAIDs: Non-Steroidal Anti-Inflammatory Drugs.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-022-03745-5>.

Additional file 1: Figure S1. Demographic distribution on Penn Medicine EHR on left and Penn Medicine Biobank on the right. **Table S1.** Drugs queried for the study. **Table S2.** Numbers of samples with ambiguous and uncalled genotypes and phenotypes from PharmCAT. **Table S3.** *CYP2D6* phenotype counts from PharmCAT research mode (without CNV and structural variants). **Table S4.** *CYP2D6* activity score counts from PharmCAT research mode (without CNV and structural variants).

Additional file 2: Regeneron Genetics Center Banner Author List and Contribution Statements.

Acknowledgements

This work was supported by the Penn Center for Precision Medicine Accelerator Fund. Additional funding provided by K23HL143161 for Dr. Sony Tuteja and PharmCAT grant: NHGRI U24HG010862.

Author contributions

SSV conceptualized and designed the study, performed the data analysis, interpreted the data, and drafted the manuscript. KK performed the data analysis, interpreted the data, and drafted the manuscript. BL performed the research, interpreted the results, and made edits to the manuscript. GH performed the research, interpreted the results, and made edits to the manuscript. MR performed the research and made edits to the manuscript. Regeneron genetics center performed the genotyping (Additional file 2). KS performed the research, interpreted the results, and made edits to the manuscript. MWC performed the research, interpreted the results, and made substantive edits to the manuscript. SD performed the research, interpreted the results, and made edits to the manuscript. AV performed the research, interpreted the results, and made edits to the manuscript. TEK conceptualized and designed the study, obtained funding, supervised the research, interpreted the results, and made substantive edits to the manuscript. MDR conceptualized and designed the study, obtained funding, supervised the research,

interpreted the results, and made substantive edits to the manuscript. ST conceptualized and designed the study, obtained funding, supervised the research, interpreted the data, and made substantive edits to the manuscript. All authors have read and approved the final manuscript.

Funding

This work was supported by the Penn Center for Precision Medicine Accelerator Fund. Additional funding provided by K23HL143161 for Dr. Sony Tuteja and PharmCAT Grant: NHGRI U24HG010862. The funders had no role in the design, interpretation and writing of the research findings.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. Gene allele frequencies for Penn Medicine Biobank participants are accessible at this link: <https://pmbb.med.upenn.edu/allele-frequency/>.

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of the University of Pennsylvania and complied with the principles set out in the Declaration of Helsinki. All individuals recruited for the Penn Medicine BioBank (PMBB) are patients of clinical practice sites of the University of Pennsylvania Health System. Appropriate consent was obtained from each participant regarding storage of biological specimens, genetic sequencing/genotyping, and access to all available EHR data.

Consent for publication

Not applicable.

Competing interests

MDR was on the Scientific Advisory Board for CIPHEROME until June 2022. All other authors declare no competing interests.

Author details

¹Department of Pathology & Laboratory Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA. ²Genomics & Computational Biology PhD Program, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA. ³Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. ⁴Department of Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA. ⁵Department of Genetics and Institute for Biomedical Informatics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA. ⁶Department of Biomedical Data Science and Medicine (BMIR), Stanford University, Stanford, CA, USA.

Received: 15 September 2022 Accepted: 30 October 2022

Published online: 28 November 2022

References

- Relling MV, Klein TE. CPIC: clinical pharmacogenetics implementation consortium of the pharmacogenomics research network. *Clin Pharmacol Ther.* 2011;89(3):464–7.
- Relling MV, Klein TE, Gammal RS, Whirl-Carrillo M, Hoffman JM, Caudle KE. The clinical pharmacogenetics implementation consortium: 10 years later. *Clin Pharmacol Ther.* 2020;107(1):171–5.
- Caudle KE, Sangkuhl K, Whirl-Carrillo M, Swen JJ, Haidar CE, Klein TE, et al. Standardizing CYP2D6 genotype to phenotype translation: consensus recommendations from the clinical pharmacogenetics implementation consortium and dutch pharmacogenetics working group. *Clin Transl Sci.* 2020;13(1):116–24.
- Gong L, Whirl-Carrillo M, Klein TE. PharmGKB, an integrated resource of pharmacogenomic knowledge. *Curr Protoc.* 2021;1(8):e226.
- Whirl-Carrillo M, Huddart R, Gong L, Sangkuhl K, Thorn CF, Whaley R, et al. An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther.* 2021;110(3):563–72.
- Wu AH, White MJ, Oh S, Burchard E. The Hawaii clopidogrel lawsuit: the possible effect on clinical laboratory testing. *Per Med.* 2015;12(3):179–81.
- Ramirez AH, Shi Y, Schildcrout JS, Delaney JT, Xu H, Oetjens MT, et al. Predicting warfarin dosage in European-Americans and African-Americans using DNA samples linked to an electronic health record. *Pharmacogenomics.* 2012;13(4):407–18.
- Wei WQ, Denny JC. Extracting research-quality phenotypes from electronic health records to support precision medicine. *Genome Med.* 2015. <https://doi.org/10.1186/s13073-015-0166-y>.
- Mosley JD, Shaffer CM, Van Driest SL, Weeke PE, Wells QS, Karnes JH, et al. A genome-wide association study identifies variants in KCNIP4 associated with ACE inhibitor-induced cough. *Pharmacogenomics J.* 2016;16(3):231–7.
- Bush WS, Crosslin DR, Owusu-Obeng A, Wallace J, Almoguera B, Basford MA, et al. Genetic variation among 82 pharmacogenes: the PGRNseq data from the eMERGE network. *Clin Pharmacol Ther.* 2016;100(2):160–9.
- Samwald M, Xu H, Blagec K, Empey PE, Malone DC, Ahmed SM, et al. Incidence of exposure of patients in the United States to multiple drugs for which pharmacogenomic guidelines are available. *PLoS One.* 2016;11(10):e0164972.
- Chanfreau-Coffinier C, Hull LE, Lynch JA, DuVall SL, Damrauer SM, Cunningham FE, et al. Projected prevalence of actionable pharmacogenetic variants and level A drugs prescribed among us veterans health administration pharmacy users. *JAMA Netw Open.* 2019;2(6):e195345.
- Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized names for clinical drugs: RxNorm at 6 years. *J Am Med Inform Assoc.* 2011;18(4):441–8.
- Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature.* 2020;586(7831):749–56.
- Sun BB, Kurki MI, Foley CN, Mechakra A, Chen CY, Marshall E, et al. Genetic associations of protein-coding variants in human disease. *Nature.* 2022;603(7899):95–102.
- McInnes G, Lavertu A, Sangkuhl K, Klein TE, Whirl-Carrillo M, Altman RB. Pharmacogenetics at scale: an analysis of the UK Biobank. *Clin Pharmacol Ther.* 2021;109(6):1528–37.
- Blout Zawatsky CL, Shah N, Machini K, Perez E, Christensen KD, Zouk H, et al. Returning actionable genomic results in a research biobank: analytic validity, clinical implementation, and resource utilization. *Am J Hum Genet.* 2021;108(12):2224–37.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10(2):giab008.
- Sangkuhl K, Whirl-Carrillo M, Whaley RM, Woon M, Lavertu A, Altman RB, et al. Pharmacogenomics clinical annotation tool (PharmCAT). *Clin Pharmacol Ther.* 2020;107(1):203–10.
- Hicks JK, El Rouby N, Ong HH, Schildcrout JS, Ramsey LB, Shi Y, et al. Opportunity for genotype-guided prescribing among adult patients in 11 US health systems. *Clin Pharmacol Ther.* 2021;110(1):179–88.
- Peterson PE, Nicholson WT, Moyer AM, Arendt CJ, Smischney NJ, Seelhammer TG, et al. Description of pharmacogenomic testing among patients admitted to the intensive care unit after cardiovascular surgery. *J Intensive Care Med.* 2021;36(11):1281–5.
- David SP, Singh L, Pruitt J, Hensing A, Hulick P, Meltzer DO, et al. The contribution of pharmacogenetic drug interactions to 90-day hospital readmissions: preliminary results from a real-world healthcare system. *J Pers Med.* 2021;11(12):1242.
- Truong TM, Apfelbaum J, Shahul S, Anitescu M, Danahey K, Knoebel RW, et al. The ImPreSS trial: implementation of point-of-care pharmacogenomic decision support in perioperative care. *Clin Pharmacol Ther.* 2019;106(6):1179–83.
- Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, et al. A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med.* 2013;369(24):2294–303.
- Gage BF, Bass AR, Lin H, Woller SC, Stevens SM, Al-Hammadi N, et al. Effect of genotype-guided warfarin dosing on clinical events and

anticoagulation control among patients undergoing hip or knee arthroplasty: the GIFT randomized clinical trial. *JAMA*. 2017;318(12):1115–24.

26. Zhu Y, Moriarty JP, Swanson KM, Takahashi PY, Bielinski SJ, Weinshilboum R, et al. A model-based cost-effectiveness analysis of pharmacogenomic panel testing in cardiovascular disease management: preemptive, reactive, or none? *Genet Med*. 2021;23(3):461–70.
27. Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, Jenkins W, Dishman E (2019) The “All of Us” research program. *N Engl J Med* 381(7):668–76.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

