

## ARTICLE

# Pharmacogenomic landscape of Indian population using whole genomes

S. Sahana<sup>1,2</sup> | Rahul C. Bhojar<sup>1</sup> | Ambily Sivadas<sup>3</sup> | Abhinav Jain<sup>1,2</sup> |  
Mohamed Imran<sup>1,2</sup> | Mercy Rophina<sup>1,2</sup> | Vigneshwar Senthivel<sup>1,2</sup> |  
Mohit Kumar Diwakar<sup>1,2</sup> | Disha Sharma<sup>1</sup> | Anushree Mishra<sup>1</sup> |  
Sridhar Sivasubbu<sup>1,2</sup> | Vinod Scaria<sup>1,2</sup>

<sup>1</sup>CSIR Institute of Genomics and Integrative Biology, New Delhi, India

<sup>2</sup>Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

<sup>3</sup>Division of Nutrition, St. John's Research Institute, St. John's National Academy of Health Sciences, Bangalore, India

## Correspondence

Vinod Scaria and Sridhar Sivasubbu, CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), South Campus, Mathura Road, New Delhi, 110025, India.

Emails: vinods@igib.in and sridhar@igib.in; s.sivasubbu@igib.res.in

## Funding information

This study was funded by the Council of Scientific and Industrial Research, India (MLP1809 and MLP2001); CSIR fellowship (to A.J. and M.K.D.); and Intel Research Fellowship (to D.S.)

## Abstract

Ethnic differences in pharmacogenomic (PGx) variants have been well documented in literature and could significantly impact variability in response and adverse events to therapeutics. India is a large country with diverse ethnic populations of distinct genetic architecture. India's national genome sequencing initiative (IndiGen) provides a unique opportunity to explore the landscape of PGx variants using population-scale whole genome sequences. We have analyzed the IndiGen variation dataset ( $N = 1029$  genomes) along with global population scale databases to map the most prevalent clinically actionable and potentially deleterious PGx variants among Indians. Differential frequencies for the known and novel variants were studied and interaction of the disrupted PGx genes affecting drug responses were analyzed by performing a pathway analysis. We have highlighted significant differences in the allele frequencies of clinically actionable PGx variants in Indians when compared to the global populations. We identified 134 mostly common (allele frequency [AF]  $> 0.1$ ) potentially deleterious PGx variants that could alter or inhibit the function of 102 pharmacogenes in Indians. We also estimate that on, an average, each Indian individual carried eight PGx variants (single nucleotide variants) that have a direct impact on the choice of treatment or drug dosing. We have also highlighted clinically actionable PGx variants and genes for which preemptive genotyping is most recommended for the Indian population. The study has put forward the most comprehensive PGx landscape of the Indian population from whole genomes that could enable optimized drug selection and genotype-guided prescriptions for improved therapeutic outcomes and minimizing adverse events.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Clinical and Translational Science* published by Wiley Periodicals LLC on behalf of the American Society for Clinical Pharmacology and Therapeutics.

## Study Highlights

### WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

The pharmacogenomic (PGx) markers related to therapeutic response and adverse drug reactions are known to vary significantly in allelic frequency across populations. Several population-scale genome sequencing studies have attempted to identify ethnic differences in the distribution of PGx variants across the global human populations, including South Asian populations. However, the genetic diversity of the Indian population is not sufficiently represented in these studies. Previous Indian studies are limited to few drugs, variants, and samples using genotyping-based approaches.

### WHAT QUESTION DID THIS STUDY ADDRESS?

India, the second most populous country in the world, is marked by distinct genetic heterogeneity owing to its diverse culture, social, and biological behaviors. Previous studies have highlighted differences in drug response between Indians and other ethnic groups. In this study, we aim to provide an integrated knowledge on the PGx landscape of diverse Indian populations using whole genome sequences which could be a center for pharmacotherapy in the future.

### WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

A population-specific PGx landscape of Indian population has been cataloged by this study, which highlights clinically actionable PGx variants and pharmacogenes for which preemptive genotyping could be prioritized specifically for the Indian population. This rich compendium of PGx markers include rare and common PGx variants including single nucleotide variants, indels, and haplotypes, including HLA alleles.

### HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

The insights from this study can help foster future PGx validation studies specific to the Indian population, provide a new perception for clinicians in decision making, and inform national drug policy decisions toward ensuring improved therapeutic outcomes for the population.

## INTRODUCTION

Interindividual variability in therapeutic response is well-documented and genetic differences are suggested to be a significant contributor apart from a variety of other factors.<sup>1</sup> Pharmacogenomics (PGx) is an emerging approach to optimize selection and dosage of therapeutics to minimize adverse drug reactions (ADRs) and to maximize the drug efficacy.<sup>2,3</sup> Several studies have correlated the variations in genes associated with drug absorption, distribution, metabolism, excretion and toxicity with differential drug responses. The US Food and Drug Administration (FDA) has included PGx markers in the product labels for over 300 drugs and the Clinical Pharmacogenetics Implementation Consortium (CPIC) have issued dosing guidelines for more than 80 drugs and over 20 genes based on curated scientific evidence encompassing a number of specialities and indications.<sup>4,5</sup>

The term ethnicity encompasses both genetic and environmental factors with shared origins, social background,

and culture, and is different from that of the race. Ethnic differences in PGx variants are now well-documented in literature.<sup>6</sup> The PGx markers related to therapeutic response and ADRs for some molecules vary significantly in allelic frequency across populations.<sup>7</sup> The advances in next-generation sequencing technologies have catalyzed a remarkable progress in the study of population-scale genomes. Such large-scale genome sequencing projects could provide interesting insights into the architecture of PGx variants<sup>8</sup> and can potentially uncover significant differences in population-wise distribution of PGx variants.<sup>6</sup> The evidence base for clinical translation of PGx for distinct populations can be enhanced significantly by incorporating multiple global populations.<sup>9</sup> A study on genome and exome sequences of the Qatari population has found 3579 potentially deleterious PGx variants involving 1163 genes which are associated with 1565 drugs, revealing 83 highly prevalent variants associated with 629 drugs.<sup>9</sup>

India is a country with extremely diverse cultural, social, and biological behaviors marked with notable

genetic heterogeneity.<sup>10</sup> The country is considered a treasure for geneticists as it consists of more than 4500 anthropologically distinct populations representing different caste, tribe, and religious groups that differs on the bases of cultural practices, linguistics, and their genetic architecture.<sup>11</sup> Previous PGx studies have highlighted disparities in drug response between Indians and other ethnic groups.<sup>12</sup> An integrated knowledge about genetic information of diverse Indian populations will be enormous and, if utilized properly, can become the center of pharmacotherapy.<sup>12</sup> India's own national genome sequencing initiative (IndiGen) program had sequenced the whole genomes of 1029 unrelated Indian individuals representing diverse ethnic groups of India and has also mined the allele frequencies of several clinically significant genetic variants to estimate the population prevalence for diverse clinical applications.<sup>13</sup>

PGx in clinical practice, clinical trials, and pharmacovigilance is still in its infancy in India. There is a dire need for promoting clinical research and pharmacovigilance programs that can help establish population-specific therapeutic strategies. In the absence of definitive country-specific policies and PGx guidelines by Indian drug regulatory agencies, PGx guidelines issued by international consortia, such as the CPIC, and labeling for approved drugs provide the much-required foundational evidence for prioritizing drug-gene candidates for promoting population-specific PGx initiatives.

In this study, we have utilized the IndiGen variation dataset to compile a comprehensive catalog of pharmacogenomic variations in the Indian population by investigating the differential frequencies of known and potentially deleterious PGx markers in the population. We have also highlighted clinically actionable PGx variants, which are enriched among Indians that could provide a new perception for the clinicians in decision making and would also pave a way for improved therapeutic outcomes.

## MATERIALS AND METHODS

### Study population and datasets

The genetic variants and their allele frequencies in the Indian population were obtained from variation datasets in Variant Calling Format (VCF) of 1029 whole genome sequences of unrelated Indian individuals, sequenced as a part of the IndiGen study.<sup>13</sup> The variants were annotated according to the GRCh38 human reference genome and the variant file includes genotype information of 55,898,112 variations, which includes single nucleotide variations (SNVs) and indels.

### Quality control

We performed genotype and individual level missingness tests (95%) and Hardy-Weinberg disequilibrium test ( $p < 5 \times 10^{-7}$ ) using PLINK version 1.09<sup>14</sup> and obtained 53,672,515 variants.

### Pharmacogenomic Variant Analysis workflow

#### Variant annotation

ANNOVAR<sup>15</sup> was used to annotate variants using database-single-nucleotide polymorphism version 150 (dbSNP v150) and RefGene databases, and also to estimate the variants allele frequencies in 1000 Genomes Phase 3 (1KGP3-ALL),<sup>16</sup> Genome Aggregation Database (gnomAD-ALL),<sup>17</sup> and Greater Middle East (GME-ALL) variome<sup>18</sup> databases. We also estimated the allele frequencies of the subpopulation datasets, such as East Asian (EAS), South Asian (SAS), Admixed American (AMR), European (EUR), and African (AFR) populations of the above databases to compare with the Indian allele frequencies.

#### Prediction of potential deleterious variants

The functional impact of exonic variants were predicted using SIFT,<sup>19</sup> PolyPhen2,<sup>20</sup> and MutationTaster.<sup>21</sup> The exonic nonsynonymous variants that were predicted deleterious (SIFT: Damaging; PolyPhen2: Probably Damaging and MutationTaster2: Disease\_causing) by at least two of these tools were taken for downstream analysis.

### Annotation of PGx variants from PharmGKB and drugbank

Clinical Annotations of PGx variants (released on December 5, 2020) were obtained from PharmGKB database,<sup>22</sup> which included 4559 annotations linked to SNVs and haplotype variants. Clinically actionable variants with the highest level of evidence (level 1A/1B) from the above list were overlapped with the potentially deleterious variants in the Indian genomes to evaluate the prevalence of these variants in Indians. Population-specific allele frequencies of the relevant PGx variants were estimated from the IndiGen database, 1KGP3 database, gnomAD database, and GME variome database. A comprehensive list of pharmacogenes was downloaded from the DrugBank database.<sup>23</sup> The predicted deleterious variants in the Indians

were overlapped with this list to generate a list of PGx variants in the Indian population.

## Statistical analysis

Fisher's exact test was used to compare the Indian allele frequencies (IndiGen) with that of global populations (1KGP3-ALL and gnomAD-ALL) to assess the frequency differences in the Indian population.

## Annotation of pharmacogenomic haplotype variants

Toward generating an exhaustive compendium of PGx associated haplotyping results, we used three different tools to call the haplotype variants from pharmacogenes from IndiGen and the 1KGP3 datasets. (i) Stargazer,<sup>24</sup> a tool for genotyping more than 50 PGx genes from next generation sequencing data, was used to call the star alleles in the PGx genes from the whole genome variation data (VCF file) of the IndiGen and 1KGP3 databases. (ii) Cyrius,<sup>25</sup> a software tool that accurately genotypes CYP2D6 was used to call the star alleles in the CYP2D6 gene from the whole genome data (bam file) of the IndiGen and 1KGP3 database. (iii) xHLA,<sup>26</sup> an algorithm for HLA typing that refines the mapping results at the amino acid level and accurately generates four-digit typing for both class I and II HLA genes was used to type PGx associated HLA haplotypes. Then, allele frequencies (AFs) of these star alleles were calculated using a python script to evaluate the prevalence of these variants in the Indian population and the 1KGP3 database.

## Construction of drug pathways and visualization

Pharmacogenes harboring PGx variants that are functionally disrupted in the Indian population with an allele frequency of more than 10% (IndiGen: 0.1) were fetched. A Sankey diagram, depicting the drug function disruption pathway, was generated using flourish studio<sup>27</sup> by mapping these genes to the associated drugs in the DrugBank database.

## RESULTS

### Summary of variants and variant annotations

The genetic variation dataset utilized for this study encompassed whole genomes obtained from 1029 unrelated Indian individuals (IndiGen data) aligned to the latest GRCh38 human reference genome. After rigorous quality control, we used 53,672,515 variants, which included SNVs and indels for our study. Genomic annotation using NCBI RefSeq identified 36,998 variants to cause nonsynonymous amino acid substitutions, which also included variants leading to gain or loss of stop codon. A total of 15,917 variants were predicted to be deleterious by at least two of the computational prediction tools and used for further downstream analysis.

### Clinically actionable pharmacogenomic variants among Indians

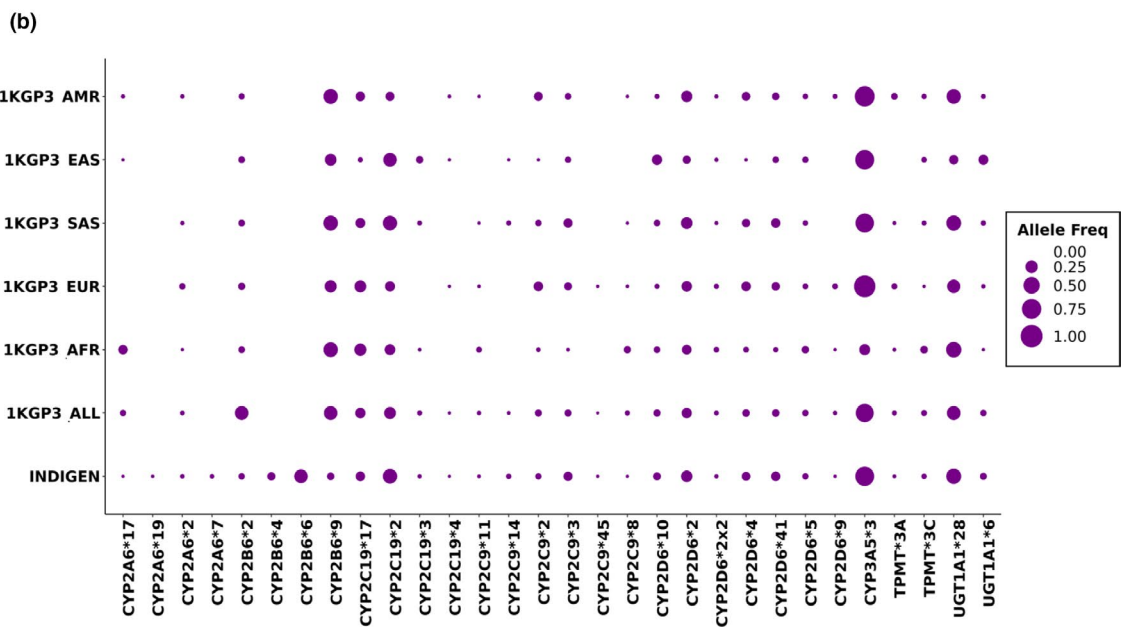
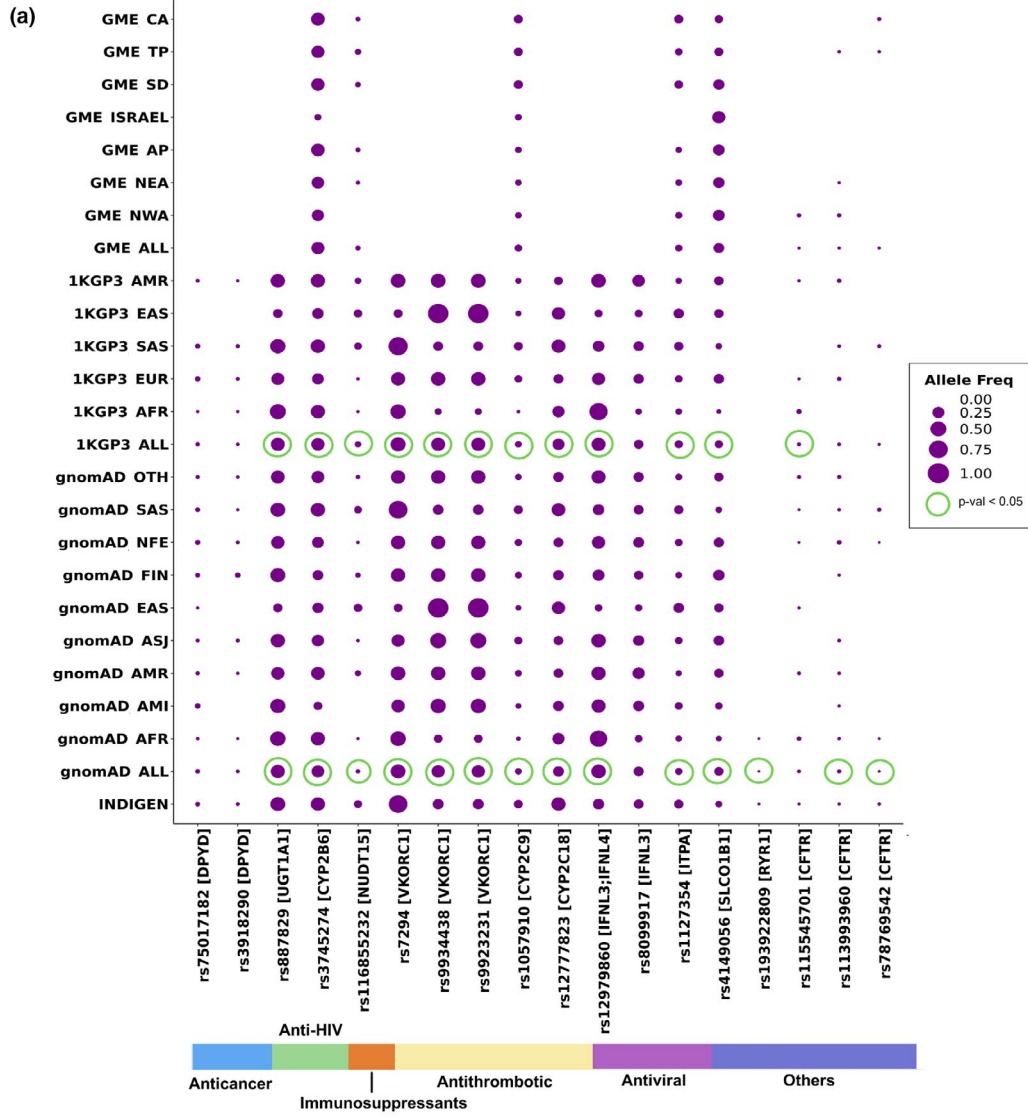
#### Single nucleotide variants

Of the 4559 clinical annotations in the PharmGKB database, 114 SNVs and 69 haplotype variants had the highest level of evidence (level 1A/1B). These variants were then overlapped with the variants in Indian population and their prevalence was evaluated based on their allele frequencies in Indians and other global population datasets, which included gnomAD, 1KGP3, and GME to indicate the remarkable interpopulation differences (Table S1).

Eighteen of the 114 clinically actionable PGx SNVs and 34 haplotype variants associated with 67 clinical annotations were found in Indian population (Figure 1a,b). We derived the allele frequencies of these variants in the IndiGen database, 1KGP3 database, gnomAD database, and Greater Middle East (GME) variome database. Statistical analysis revealed 14 SNVs whose allele frequencies in Indians are significantly different ( $p$  value  $< 0.05$ ) from the global population (gnomAD-ALL and 1KGP3-ALL) averages (Figure 1a).

Several important differences in allele frequency were identified. For example, *SLCO1B1* variant rs4149056 associated with simvastatin toxicity occurs at a considerably

**FIGURE 1** (a) Comparison of Indian allele frequencies of clinically relevant PGx single nucleotide variants with populations in 1KGP3 dataset, gnomAD database, and GME database. PGx variants in Indians which yielded significant  $p$  value ( $p < 0.05$ ) in Fisher's exact test comparing Indian allele frequency with other databases are highlighted in green outer circle. (b) Comparison of Indian allele frequencies of clinically relevant PGx haplotype variants with populations in the 1KGP3 dataset. AFR, African/African American; AMI, Amish; AMR, Admixed American/Latino; ASJ, Ashkenazi Jewish; CA, Central Asia; EAS, East Asian; EUR, European; FIN, Finnish; GME, Greater Middle East; gnomAD, Genome Aggregation Database; IKGP3, 1000 Genomes Phase 3; NEA, Northeast Africa; NWA, Northwest Africa; OTH, Other (population not assigned); PGx, pharmacogenomic; SAS, South Asian; SD, Syrian Desert; TP, Turkish Peninsula



lower frequency (IndiGen: 0.05) compared to the global populations (gnomAD-ALL: 0.12). Notably, three variants in the *VKORC1* gene (rs9923231 [IndiGen: 0.18], rs9934438 [IndiGen: 0.18], and rs7294 [IndiGen: 0.71]) associated with warfarin dosage and efficacy was found to be most common among Indians. The prevalence of rs7294 variant in the South Asian population is significantly higher when compared to the global populations (IndiGen: 0.71; gnomAD-SAS: 0.72; 1KGP3-SAS: 0.75; gnomAD-ALL: 0.40; and 1KGP3-ALL: 0.42).

We also observed that the *DPYD* variant rs3918290 associated with fluoropyrimidine-based chemotherapy drug toxicity showed increased prevalence in India (IndiGen: 0.005) among the South Asian populations (gnomAD SAS: 0.001). Similarly, *NUDT15* variant rs116855232 associated with azathioprine and mercaptopurine dosage and toxicity showed highest prevalence in Asian populations (IndiGen: 0.08; gnomAD-SAS: 0.07; 1KGP3-SAS: 0.07; gnomAD-EAS: 0.1; and 1KGP3-EAS: 0.09) and lowest in African populations (gnomAD-AFR: 0.001; and 1KGP3-AFR: 0.0008). The intronic variant rs12979860 associated with the efficacy of peginterferon based regimens showed a significantly different prevalence among the Asian population, i.e., the prevalence was 20% in the South Asian populations, including Indians (IndiGen: 0.20) whereas it was less than 10% in the East Asian populations.

## Haplotype variants

*CYP2B6* haplotypes comprising *CYP2B6\*2* (IndiGen: 0.03), *CYP2B6\*4* (IndiGen: 0.07), *CYP2B6\*6* (IndiGen: 0.31), and *CYP2B6\*9* (IndiGen: 0.05) that are associated with the metabolism/dosage/toxicity of antiretroviral medication efavirenz are found to be prevalent among Indians. This suggests that a striking 60% of Indians require a lower dose of efavirenz as they carry at least one copy of the decreased function allele *CYP2B6\*6* or *CYP2B6\*9* as per the CPIC dosing guidelines.<sup>28</sup> The prevalence of the variant *UGT1A1\*28* associated with the toxicity of the antiretroviral drug atazanavir/ritonavir and antineoplastic drugs (FOLFIRI/irinotecan) is higher in Indians compared to the global population (IndiGen: 0.40; and 1KGP3-ALL: 0.32). We have also noticed that the prevalence of the variant *UGT1A1\*6* associated toxicity of the antineoplastic drugs (FOLFIRI/irinotecan) is significantly higher among East Asians in comparison to the South Asian populations (IndiGen: 0.05; 1KGP3-SAS: 0.02; and 1KGP3-EAS: 0.1).

In comparison to global populations, *CYP2C19\*2* variant associated with metabolism or efficacy of one antiplatelet drug, three proton-pump inhibitor drugs, one antifungal drug, and seven antidepressant drugs is found in higher allele frequency in Indians (IndiGen: 0.36; and

1KGP3-ALL: 0.22; Table S2). This translates to key actionable guideline recommendations in 13% of Indian patients who are *CYP2C19* poor metabolizers (\*2/\*2), such as alternate antifungal therapy in case of voriconazole, dosage reduction for citalopram and escitalopram, alternate antiplatelet therapy in case of clopidogrel, and dosage reduction for the selective serotonin reuptake inhibitor sertraline.

The variant *CYP3A5\*3* associated with dosing of the commonly used immunosuppressant drug tacrolimus is less prevalent in African populations who are therefore known to require higher tacrolimus dosage compared to all the other populations (IndiGen: 0.7; 1KGP3-AFR: 0.18; 1KGP3-EUR: 0.94; 1KGP3-SAS: 0.66; 1KGP3-EAS: 0.71; and 1KGP3-AMR: 0.79). We also observed that *TPMT\*3A* and *TPMT\*3C* variants associated with the dosage reduction and toxicity of immunosuppressant drugs, such as azathioprine and mercaptopurine, is found to be less prevalent in Indians (IndiGen: 0.003 and 0.018) compared to the global population (1KGP3-ALL: 0.012 and 0.026), respectively.

The variant *CYP2C9\*3* is associated with the metabolism and response to NSAIDs, such as the widely prescribed drug ibuprofen. The prevalence of this variant is 10% (IndiGen: 0.1) among Indians and South Asian populations. As per the CPIC guidelines,<sup>29</sup> our analysis shows that at least 10% of Indians carrying the variant would be possibly poor or intermediate metabolizers thereby requiring a lower dose.

About 25% of clinically used drugs, such as antidepressants, antipsychotics, and opioids, are mostly metabolized by the *CYP2D6* enzyme and their activity is highly variable from poor metabolism to ultrarapid metabolism.<sup>30</sup> There are around seven clinically actionable *CYP2D6* haplotype variants present in the Indian population associated with two antipsychotics, four opioids, and nine antidepressants (Table S2). The reduced functional allele *CYP2D6\*41* was found to be the most common allele among the Indians and is associated with the intermediate metabolizer phenotype.<sup>31</sup> Our analysis confirms that the prevalence of this variant is higher among Indians (IndiGen: 0.11; and 1KGP3-ALL: 0.06).

## HLA haplotype variants

The prevalence of *HLA-A\*31:01:02* (IndiGen: 0.2) and *HLA-B\*15:02:01* (IndiGen: 0.3) variants associated with severe cutaneous adverse drug reactions (SCARs) like Stevens-Johnson syndrome of anticonvulsant drugs, carbamazepine, oxcarbazepine, and phenytoin (Table S2) are 2% and 3% in the overall Indian population, respectively. Similarly, the *HLA-B\*58:01* variant associated with

the toxicity of allopurinol, a xanthine reductase inhibitor drug, is found to be prevalent in 3.7% of the Indian population (IndiGen: 0.37). The variant *HLA-B\*57:01:01* associated with the hypersensitivity to abacavir, an antiretroviral drug found to be present in 4% of the overall Indian population (IndiGen: 0.4). However, the allele frequencies of these variants seem to be immensely higher in the South Indian states Kerala, Andhra Pradesh, Karnataka, and Tamil Nadu, as per the allele frequencies given in the Allele Frequency Net Database.<sup>32</sup>

## Actionable pharmacogenomic variants in individual genomes

We analyzed the carrier status of clinically actionable variants (PharmGKB level 1A/1B) in each of the 1029 genomes. Of the 18 SNVs and 34 haplotypes identified in the Indian population, we observed that, on average, each Indian individual carries eight actionable PGx variants (SNVs) that have a direct impact on the choice of treatment or drug dosing.

## Potentially deleterious variants in pharmacogenes among Indians

To capture a comprehensive overview of the extent of pharmacogene disruption among Indians, we analyzed overlap between the putative deleterious variants in IndiGen data with 2411 pharmacogenes associated with 4015 drugs, as listed in DrugBank. The analysis revealed 14,752 PGx variants, including 4366 novel variants (dbSNP v150) potentially hampering the function of 2184 genes, including 145 transporter/carriers, 156 enzymes, and 984 targets associated with 1740 drugs across therapeutic areas. We also estimated the allele frequencies of these variants in the Indian population and the global population to compare their prevalences. We observed that 14,078 of 14,752 variants were rare ( $AF < 0.01$ ) and 218 variants among them were common ( $AF > 0.05$ ). We then carried out our analysis with the most common 134 variants which show an AF of more than 10% among Indians (Figure 2, Table S3).

We identified nine common variants ( $AF > 10\%$ ) among Indians that are rare ( $AF < 1\%$ ) in other global populations. The variant rs116201358 in *C8A* gene is found to be strikingly higher in frequency among South Asian populations in comparison with other subpopulations (IndiGen: 0.13; gnomAD-ALL: 0.007; and gnomAD-SAS: 0.12) and is associated with the better treatment outcome of imatinib mesylate, a drug used to treat chronic myeloid leukemia.<sup>33</sup> Similarly, another potentially deleterious

variant rs11568367 in the *ABCB11* gene showing high prevalence in the Indian population (IndiGen: 0.11; and gnomAD-ALL: 0.003) is associated with significantly impaired function of taurocholate transport.<sup>34</sup> The variant rs202242769 in the *CYP21A2* gene is found to be extremely higher in the Indian population in comparison with all the other subpopulations, including other South Asian populations (IndiGen: 0.227; gnomAD-ALL: 0.0053; and gnomAD-SAS: 0.0221) and is associated with the change in hormone levels in the patients with congenital adrenal hyperplasia as well as in some healthy individuals.<sup>35</sup> The variants in *GLRA4* gene are associated with a negative response to antipsychotic treatment in the patients with schizophrenia. We have found a potentially deleterious variant rs182137906 in the *GLRA4* gene with the higher allele frequency among Indian population and could not be found in some of the other populations (IndiGen: 0.114; gnomAD-ALL: 0.0034; gnomAD-AMI: 0; and gnomAD-EAS: 0). The variants in the *HLA-B* gene are associated with adverse reactions to a wide range of pharmaceuticals and are also susceptible and resistant to a number of diseases. A potentially deleterious variant rs1050538 in the *HLA-B* gene, which is associated with the severity of human respiratory syncytial virus infection and its diagnostics, and it is found to be highly prevalent among Indians in comparison with global populations (IndiGen: 0.159; and gnomAD: 0.0059).<sup>36</sup> We also found four potentially deleterious variants in three HLA class II genes, which occur very rarely in global populations compared to the Indian population: *HLA-DQB1* variant rs9274387 (IndiGen: 0.16; and gnomAD-ALL: 0.000009); *HLA-DQB1* variant rs9274395 (IndiGen: 0.17; and gnomAD-ALL: 0.0002); *HLA-DRB5* variant rs147669022 (IndiGen: 0.32; and gnomAD-ALL: 0.004); and *HLA-DQA2* variant rs79517313 (IndiGen: 0.12; and gnomAD-ALL: 0.0076).

In addition to these, we also observed that 14 of 134 variants are associated with 64 clinical annotations in PharmGKB irrespective of the levels of evidence. The rs1799971 variant in the *OPRM1* gene associated with the number of opioids dosage/efficacy/toxicity/ADR (level 3) was highly prevalent among Indians (IndiGen: 0.44; and gnomAD ALL: 0.12). Interestingly, the *MTHFR* variant rs1801133 associated with antineoplastic drugs, such as methotrexate, carboplatin, and cyclo-phosphamide (level 2A), was less prevalent in Indians (IndiGen: 0.145), which had global average frequency of 27%. But, the *FLT3* variant rs1933437, which is associated with toxicity of anticancer drug sunitinib (level 3), was found to have increased prevalence among Indians (IndiGen: 0.675; and gnomAD ALL: 0.5251). In addition, the *ATP7A* variant rs2227291, associated with toxicity of chemotherapy drugs docetaxel, thalidomide (level 3) was found to be in higher allele frequency in Indians (IndiGen: 0.361; and gnomAD

ALL: 0.2468). The *EPHX1* variant rs1051740 associated with dosage or efficacy of anticonvulsant drug carbamazepine (level 2B) is comparatively in higher allele frequency in India (IndiGen: 0.366; and gnomAD ALL: 0.2735). Alternatively, *PON1* variant rs854560 was less prevalent in Indians (IndiGen: 0.188; and gnomAD ALL: 0.2889), which was associated with efficacy and toxicity of the antiplatelet drug clopidogrel (level 4).

## Drug pathway analysis of disrupted PGx genes in the Indian population

In order to understand the interaction of the disrupted PGx genes affecting drug responses among Indians, we performed a pathway analysis of 708 drugs associated with 118 genes that have been disrupted in at least 10% of the population (Figure 3, Table S4). The pathway map has been represented as a Sankey diagram, which provides an overview of the drug pathway disruption categorized into different levels of gene function, such as drug transport/carriers, enzymes, and targets. We observed that 182 drugs show a minimum of 50% function disruption of its associated genes, which is involved in transport/carrier, enzymes, or targets. Of these, 91 drugs show complete disruption of at least one of the functions: transport, metabolism, or targeting. For example, an antineoplastic drug trastuzumab with the sole target gene *ERBB2* has a potentially nonfunctional variant in 65% of the Indians and it is reported to be associated with the number of serious cardiac function adverse reactions.<sup>37</sup> Of the 118 disrupted genes, four genes (*CYP2D6*, *CYP4F2*, *HLA-A*, and *HLA-B*) were associated with 22 drugs that are mentioned in the FDA table of PGx biomarkers in drug labeling.

## DISCUSSION

To date, several population-scale genome sequencing studies have attempted to characterize the ethnic disparities in the distribution of PGx variants across different global human populations, including South Asian populations. However, these studies had a limited sampling

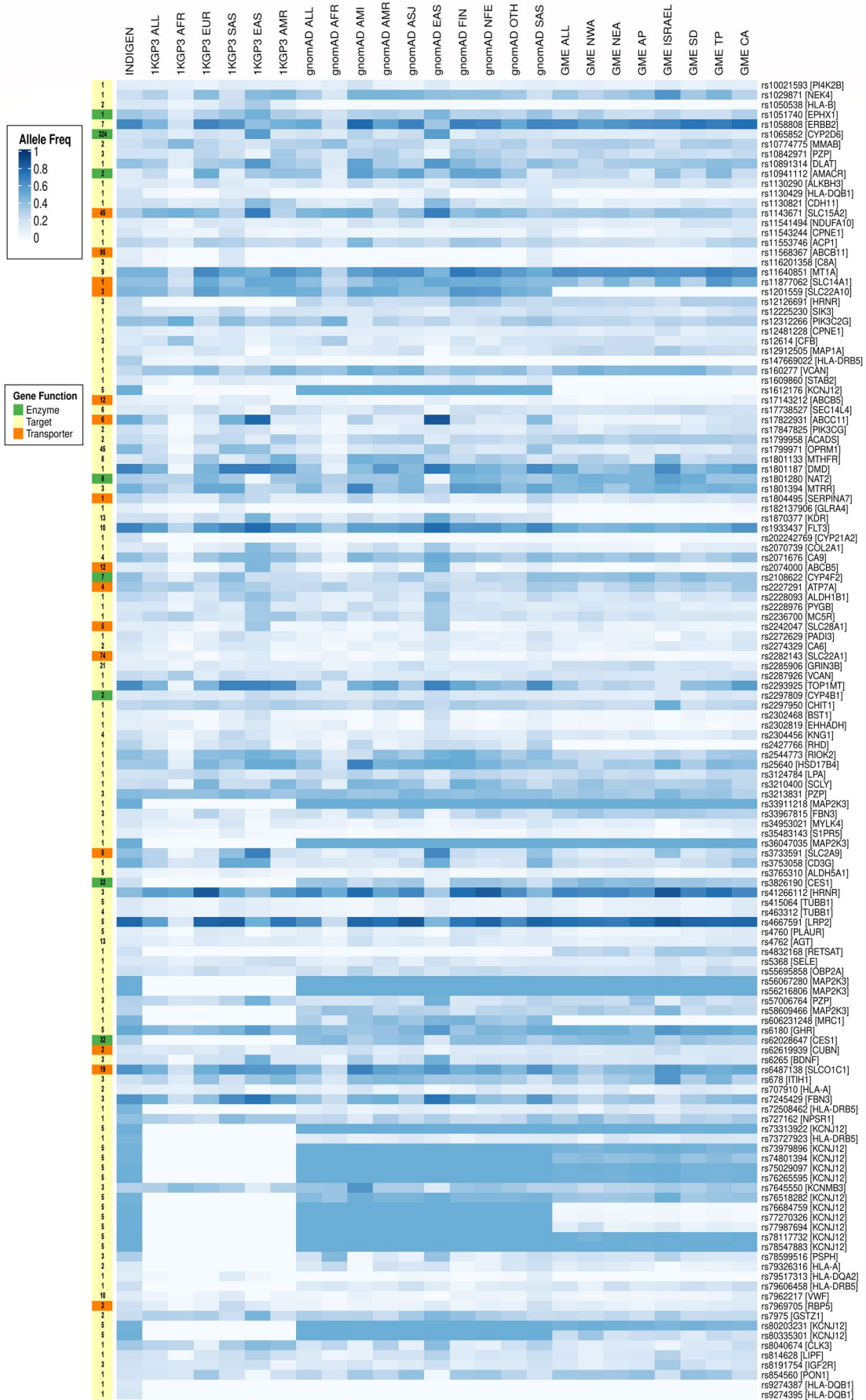
of the remarkable genetic diversity represented by the Indian population. Most of these studies also failed to comprehensively characterize PGx-associated SNVs and indels along with haplotype variants using a systematic approach. Previous genotyping-based studies have highlighted drastic differences in the prevalence of the selected PGx variants among Indians associated with widely prescribed drugs like warfarin and clopidogrel.<sup>38,39</sup>

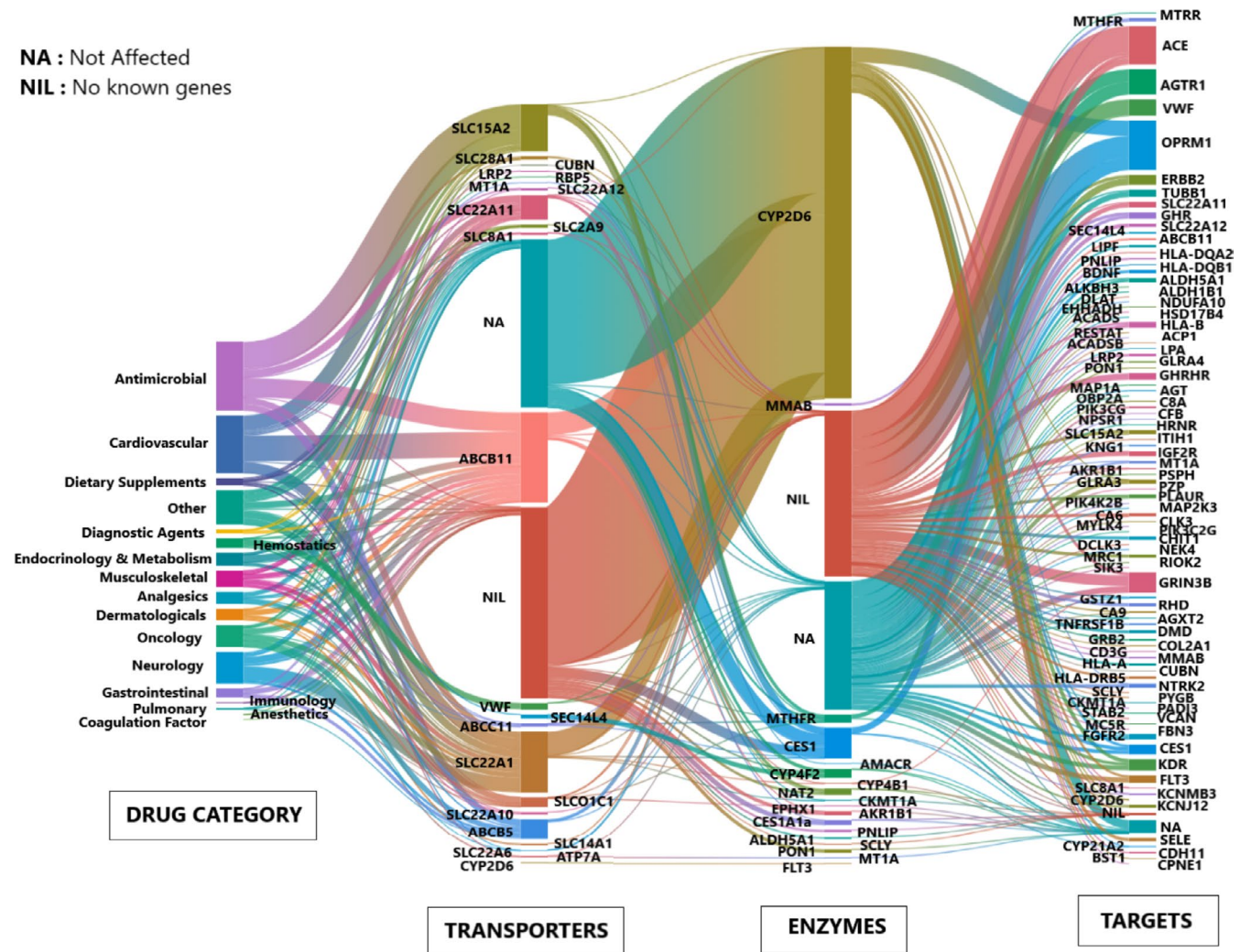
In the present study, we have analyzed large-scale Indian population genomic data to identify the most prevalent clinically actionable and potentially deleterious PGx variants among Indians. We established the presence of 18 SNVs and 34 haplotype variants, including *HLA* alleles associated with 85 clinical annotations among Indians, which has relevant guidelines for drug dosing or ADRs. We have also estimated the AF of these variants in Indians (IndiGen database) and other global population datasets and their superpopulations (gnomAD and 1KGP3 databases). Three variants in the *VKORC1* gene (rs9923231, rs9934438, and rs7294) were most common among Indians and found to be well studied to alter the warfarin pharmacodynamics.<sup>40</sup> These polymorphisms can significantly alter the warfarin pharmacodynamics and hence dose maintenance is required.<sup>41</sup> This seems to be great evidence for incorporating *VKORC1* genotype information to select the optimal dose for the individual patient at the start of the warfarin therapy in India. Of the seven actionable *CYP2D6* haplotype variants present in the Indian population, four variants were found to be in comparatively higher allele frequency with the global population. Of these, the reduced function allele *CYP2D6\*41* has remarkably higher prevalence among Indians and is associated with a number of commonly prescribed antipsychotics, opioids, and antidepressant medications.

Cystic fibrosis was initially considered to be nonexistent in India, but later reports suggested that cystic fibrosis occurs in India but its precise magnitude is not known.<sup>42</sup> It has also been reported that the frequency of the common cystic fibrosis mutation rs113993960 in the *CFTR* gene is lesser in the Indian population compared with other populations.<sup>43</sup> Our analysis also revealed a combined carrier frequency of 0.7% involving three variants in the *CFTR* gene associated with cystic fibrosis disease and its therapeutic efficacy (rs78769542, rs115545701, and

**FIGURE 2** Allele frequencies of most common potentially deleterious nonsynonymous variants (AF > 0.1) in Indians involved drug transport, metabolism and target. Allele frequencies are compared with 1KGP3 database, gnomAD database, and GME database. Y-axis represents the variant [gene name] and X-axis represents the population and subpopulation. Gene function category is color coded on the left with the number of drugs associated with each gene. AF, allele frequency; AFR, African/African American; AML, Amish; AMR, Admixed American/Latino; AP, Arabian Peninsula; ASJ, Ashkenazi Jewish; CA, Central Asia; EAS, East Asian; EUR, European; FIN, Finnish; GME, Greater Middle East; gnomAD, Genome Aggregation Database; 1KGP3, 1000 Genomes Phase 3; NEA, Northeast Africa; NFE, Non-Finnish European; NWA, Northwest Africa; OTH, Other (population not assigned); PGx, pharmacogenomic; SAS, South Asian; SD, Syrian Desert; TP, Turkish Peninsula







**FIGURE 3** Sankey diagram. Drug pathway map representing the pharmacogenes functionally disrupted in more than 10 percent of the Indian population. The first column represents the broad drug category associated with the potentially deleterious variants in the Indian population. The second, third, and fourth columns represent the pharmacogenes belonging to the classes: transporter/carriers, enzymes, and targets, respectively. The link width represents the degree of functional loss of the given drug in terms of the pharmacogene classes

rs113993960) in the Indian population in concordance with earlier findings.

We have also analyzed the clinically actionable PGx variants in individual genomes and observed that, on average, each Indian individual carries eight actionable PGx variants that could impact the choice of treatment or drug dosing in many circumstances. Additionally, our study identified 14,752 potentially deleterious PGx variants, including 4366 novel variants that could potentially alter or inhibit the function of 2184 pharmacogenes in Indians. Our analysis revealed 134 variants and categorized them as the most common variants among Indians as they show AF of more than 10%. These included nine variants which are more prevalent among Indians (AF > 10%) and rare among the other global populations (AF < 1%). We also observed that 14 of 134 variants are associated with 64 clinical annotations in the PharmGKB database irrespective of their levels of evidence.

*HLA* genes are well known to have associations with various types of SCARs when exposed to certain drug treatments.<sup>44</sup> We discovered four potentially deleterious variants in three *HLA* class II genes that were widely prevalent in Indians and not in other global populations. *HLA* alleles are standard biomarkers for abacavir, carbamazepine, and allopurinol and are associated with ADRs.<sup>45</sup> Our analysis revealed increased prevalence of three clinically actionable *HLA* haplotype variants and one potentially deleterious variant in the *HLA-B* gene that are associated with these drugs among Indians. Genotyping of *HLA* alleles prior to the phenytoin, abacavir, carbamazepine, and allopurinol therapies could prevent toxicity and improve patient outcomes.

Furthermore, we also analyzed the drug categories associated with the most commonly disrupted PGx genes in the Indian population by performing a pathway analysis. We observed 91 drugs that show complete disruption in its

transport, metabolism, or targeting functions. This analysis has revealed 22 drugs that were associated with four of the FDA specified PGx biomarkers (*CYP2D6*, *CYP4F2*, *HLA-A*, and *HLA-B*) for dosage and administration. The Sankey analysis can potentially enable selection of alternate drugs with fewer population-level pathway disruptions.

There are a lot of highly differentiated SNPs, indels, and haplotypes which have not been assigned a high level of evidence for association with drug response or toxicity. PGx validation studies for these variants are required for the implication of these variations. Implementing dosing guidelines based on variation in the PGx genes in Indian population would lead to positive outcomes.<sup>46</sup>

In conclusion, we have put forward the population specific PGx landscape of Indian population by investigating differential frequencies of PGx markers in the genomic data. We have highlighted clinically actionable PGx variants and the genes for which preemptive genotyping should be specifically recommended in Indians. We foresee that this study could provide a rich resource for designing future validation studies and enabling improved therapeutic outcomes by providing enhanced drug selection and dosing guidelines.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge Mukta Poojary, Mohit Mangla, and Arvinden VR for their consecutive support and suggestions throughout the study. The authors would like to thank Seung-been Lee for his timely help by providing the updated version of Stargazer that supports GRCh38.

## CONFLICT OF INTEREST

The authors declared no competing interests for this work.

## AUTHOR CONTRIBUTIONS

S.Sa. and A.S. wrote the manuscript. V.Sc. and S.Si. designed the research. S.Sa., R.C.B., A.J., M.I., M.R., V.Se., M.K.D., D.S., and A.M., performed the research. S.Sa. analyzed the data.

## REFERENCES

- Roden DM, George AL Jr. The genetic basis of variability in drug responses. *Nat Rev Drug Discov.* 2002;1:37-44.
- Weinshilboum RM, Wang L. Pharmacogenomics: precision medicine and drug response. *Mayo Clin Proc.* 2017;92(11):1711-1722.
- Cecchin E, Stocco G. Pharmacogenomics and personalized medicine. *Genes.* 2020;11:679.
- Center for Drug Evaluation & Research. Table of pharmacogenomic biomarkers. <https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling> (2020).
- Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines. <https://cpicpgx.org/guidelines>.
- Xie HG, Kim RB, Wood AJ, Stein CM. Molecular basis of ethnic differences in drug disposition and response. *Annu Rev Pharmacol Toxicol.* 2001;41:815-850.
- Ramos E, Doumatey A, Elkahoul AG, et al. Pharmacogenomics, ancestry and clinical decision making for global populations. *Pharmacogenomics J.* 2014;14:217-222.
- Rabbani B, Nakaoka H, Akhondzadeh S, Tekin M, Mahdieh N. Next generation sequencing: implications in personalized medicine and pharmacogenomics. *Mol Biosyst.* 2016;12:1818-1830.
- Sivadas A, Scaria V. Pharmacogenomic survey of Qatari populations using whole-genome and exome sequences. *Pharmacogenomics J.* 2018;18:590-600.
- Dyson T. *A Population History of India: From the First Modern People to the Present Day.* New York, NY: Oxford University Press; 2018.
- Mastana SS. Unity in diversity: an overview of the genomic anthropology of India. *Ann Hum Biol.* 2014;41:287-299.
- Shewade DG. Status of pharmacogenomics research in India during the last five years. *Proceedings of the Indian National Science Academy.* 10.16943/ptinsa/2017/49229.
- Jain A, Bhojar RC, Pandhare K, et al. IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes. *Nucleic Acids Res.* 2021;49:D1225-D1232.
- Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559-575.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38:e164.
- 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015;526:68-74.
- Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581:434-443.
- Scott EM, Halees A, Itan Y, et al. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet.* 2016;48:1071-1076.
- Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003;31:3812-3814.
- Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;76:20.
- Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods.* 2014;11:361-362.
- Hewett M. PharmGKB: the pharmacogenetics knowledge base. *Nucleic Acids Res.* 2002;30:163-165.
- Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledge-base for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008;36:D901-D906.
- Lee S-B, Wheeler MM, Thummel KE, Nickerson DA. Calling star alleles with Stargazer in 28 pharmacogenes with whole genome sequences. *Clin Pharmacol Ther.* 2019;106:1328-1337.
- Chen X, Shen F, Gonzaludo N, et al. Cyrius: accurate CYP2D6 genotyping using whole-genome sequencing data. *Pharmacogenomics J.* 2021;21(2):251-261.
- Xie C, Yeo ZX, Wong M, et al. Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. *Proc Natl Acad Sci USA.* 2017;114:8059-8064.
- Flourish. <https://flourish.studio/>.

28. Desta Z, Gammal RS, Gong LI, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2B6 and efavirenz-containing antiretroviral therapy. *Clin Pharmacol Ther.* 2019;106:726-733.
29. Theken KN, Lee CR, Gong L, et al. Clinical Pharmacogenetics Implementation Consortium Guideline (CPIC) for CYP2C9 and nonsteroidal anti-inflammatory drugs. *Clin Pharmacol Ther.* 2020;108:191-200.
30. Gaedigk A. Complexities of CYP2D6 gene analysis and interpretation. *Int Rev Psychiatry.* 2013;25:534-553.
31. Manoharan A, Shewade DG, Ravindranath PA, et al. Resequencing gene in Indian population: identified as the major reduced function allele. *Pharmacogenomics.* 2019;20:719-729.
32. Gonzalez-Galarza FF, McCabe A, Dos Santos EJM, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* 2020;48:D783-D788.
33. Shokeen Y, Sharma NR, Vats A, et al. Identification of prognostic and susceptibility markers in chronic myeloid leukemia using next generation sequencing. *Ethiop J Health Sci.* 2018;28:135-146.
34. Ho RH, Leake BF, Kilkenny DM, et al. Polymorphic variants in the human bile salt export pump (BSEP; ABCB11): functional characterization and interindividual variability. *Pharmacogenet Genomics.* 2010;20:45-57.
35. Doleschall M, Szabó JA, Pázmándi J, et al. Common genetic variants of the human steroid 21-hydroxylase gene (CYP21A2) are related to differences in circulating hormone levels. *PLoS One.* 2014;9:e107244.
36. Daniel TODT, Haid S, Wetzke M, et al. Diagnostics and therapy for human respiratory syncytial virus. US Patent (2020).
37. McKeage K, Perry CM. Trastuzumab. *Drugs.* 2002;62:209-243.
38. Giri AK, Khan NM, Grover S, et al. Genetic epidemiology of pharmacogenetic variations in CYP2C9, CYP4F2 and VKORC1 genes associated with warfarin dosage in the Indian population. *Pharmacogenomics.* 2014;15:1337-1354.
39. Giri AK, Khan NM, Basu A, Tandon N, Scaria V, Bharadwaj D. Pharmacogenetic landscape of clopidogrel in north Indians suggest distinct interpopulation differences in allele frequencies. *Pharmacogenomics.* 2014;15:643-653.
40. Owen RP, Gong L, Sagreiya H, Klein TE, Altman RB. VKORC1 pharmacogenomics summary. *Pharmacogenet Genomics.* 2010;20:642-644.
41. Krishna Kumar D, Shewade DG, Loriot M-A, et al. Effect of CYP2C9, VKORC1, CYP4F2 and GGCX genetic variants on warfarin maintenance dose and explicating a new pharmacogenetic algorithm in South Indian population. *Eur J Clin Pharmacol.* 2014;70:47-56.
42. Kabra SK, Kabra M, Lodha R, Shastri S. Cystic fibrosis in India. *Pediatr Pulmonol.* 2007;42:1087-1094.
43. Ashavaid TF, Raghavan R, Dhairyawan P, Bhawalkar S. Cystic fibrosis in India: a systematic review. *J Assoc Physicians India.* 2012;60:39-41.
44. Chang C-J, Chen C-B, Hung S-I, Ji C, Chung W-H. Pharmacogenetic testing for prevention of severe cutaneous adverse drug reactions. *Front Pharmacol.* 2020;11:969.
45. Zhou Y, Krebs K, Milani L, Lauschke VM. Global frequencies of clinically important HLA alleles and their implications for the cost-effectiveness of preemptive pharmacogenetic testing. *Clin Pharmacol Ther.* 2021;109:160-174.
46. Prasad N, Jaiswal A, Behera MR, et al. Melding pharmacogenomic effect of and CYP3A5 gene polymorphism on tacrolimus dosing in renal transplant recipients in Northern India. *Kidney Int Rep.* 2020;5:28-38.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Sahana S, Bhoyar RC, Sivadas A, et al. Pharmacogenomic landscape of Indian population using whole genomes. *Clin Transl Sci.* 2022;15:866–877. <https://doi.org/10.1111/cts.13153>