Faculty of Medicine with the Division of Dentistry
and Division of Medical Education in English
Medical University of Bialystok

Complete Clinical Pharmacogenomic Profiling,
new insight for the future of personalized drug therapy

By:
Alireza Tafazoli

Ph.D. dissertation as a collection of papers
in the field of medical and health sciences
in the discipline of medical sciences

Supervisor:
Prof. Wojciech Miltyk
Department of Analysis and Bioanalysis of Medicines
Medical University of Bialystok, Bialystok, Poland

Co-Supervisor:
Dr. Jesse J. Swen
Department of Clinical Pharmacy and Toxicology, Leiden University
Medical Center (LUMC), Leiden, Netherlands

Bialystok 2022

The project in current dissertation is conducted and performed at:

- Department of Analysis and Bioanalysis of Medicines,
  Faculty of Pharmacy with the Division of Laboratory Medicine
  Medical University of Bialystok, Bialystok, Poland

- Genomics Lab, Clinical Research Center,
  University Clinical Hospital, Bialystok
  Medical University of Bialystok, Bialystok, Poland

- Department of Clinical Pharmacy and Toxicology,
  Leiden University Medical Center, Leiden, Netherlands

Funding:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754432 and the Polish Ministry of Science and Higher Education, from financial resources for science in 2018-2023 granted for the implementation of an international co-financed project.

Dedicated to:

- *My Wife,* ***Zahra****, who has ignored her dreams to support my dreams.*
- *My Son,* ***Arya****, to know the beauty and pleasure of discovery in science.*
- *My Professor,* ***Wojciech Miltyk****, who was the kindest and most supportive promoter I've ever had during my academic career.*
- *My Co-supervisor,* ***Jesse J. Swen****, who provided me a huge amount of help to manage and finalize my Ph.D. project.*

Acknowledgement:

- *I would like to thank* ***Dr. Maaike van der Lee*** *for her invaluable toturials during my stay in Netherlands.*
- *My deepest thank to* ***Prof. George P. Patrinos*** *and* ***Prof. Mohammad Reza Abbaszadegan and Prof. Barbara Bojko*** *for reviewing this dissertation and providing precious comments on that.*

“If I have seen further it is by standing on the shoulders of Giants”.

Sir Isaac Newton

“If you can't explain it simply, you don't understand it well enough”.

Albert Einstein

Table of Contents

1. List of articles included in the dissertation

- Original article:

- **Tafazoli, A.**, van der Lee, M., Swen, J., Zeller, A., Wawrusiewicz-Kurylonek, N., Mei, H., Vorderman, R., Konopko, K., Zankiewicz, A., and Miltyk, W., (2022). Development of an extensive workflow for comprehensive clinical Pharmacogenomic profiling: *lessons from a pilot study on 100 Whole Exome Sequencing data. Submitted manuscript*. *Pharmacogenomics J*. DOI: 10.1038/s41397-022-00286-4. IF: 3.550, MeiN: 140

- Review article:

- **Tafazoli, A.**, Guchelaar, H.J., Kretowski, A., Miltyk, W, Swen, J.J. (2021). Applying next-generation sequencing platforms for Pharmacogenomics testing in clinical practice, *Front. Pharmacol, Pharmacogenetics and Pharmacogenomics section*. DOI: 10.3389/fphar.2021. (12) – 693453. IF: 5.988, MEiN: 100
- **Tafazoli, A.**, Kurylonek, N.W., Posmyk, R., Miltyk, W. (2021). Pharmacogenomics, how to deal with different types of variants in next generation sequencing data in the personalized medicine area, *J. Clin. Med*. Doi: 10.3390/jcm10010034. IF: 4.964, MEiN: 140

2. List of a candidate's publications:

(Number of a publications, Impact Factor, points of Ministry of Science and Higher Education (MNiSW)):

- Articles included in the dissertation
- Articles not included in the dissertation
- Conference abstracts

Article type	**Number**	**Impact Factor**	**MNiSW points**
Articles included in the dissertation	3	14.51	380
Articles not included in the dissertation	20	57.80	1410
Conference abstracts	1	-	-
Summary	24	72.31	1790

3. Introduction:

Pharmacogenetics, Pharmacogenomics, and Personalized Medicine

There has been significant growth in using the genomic profile for personalized clinical care in recent years (1). Pharmacogenetics, as one of the main aspects of this topic, evaluates the functional genetic variations in people, which are responsible for mechanisms of drug responses to certain medications, possible adverse drug reactions (ADRs), drug toxicity, efficiency, etc. In this field, Pharmacogenomics (PGx) works with whole-genome or a large set of genes, related to the therapeutic issues, simultaneously in a more comprehensive analysis (2). People with particular variants in drug-related genes cannot metabolize medicines properly and thus become vulnerable to ADRs. Investigations indicated that variants in a certain gene or a group of genes may determine the observed effect of a particular drug (3). Although, particular set of genes determine the drug response in individuals, certain genes are more likely to be involved in the absorption, distribution, metabolism, and excretion (ADME) of a wide variety of drugs. A number of 34 genes has been introduced as the core ADME genes. Additionally, extended ADME gene list includes hundreds of additional genes for proteins responsible for the modification of functional groups of drugs, conjugation of drugs with internal components, the uptake or excretion of drugs in and out of cells, and those that may alter the expression of other ADME genes or affect the biochemistry of ADME enzymes. Moreover, up to 800 genes have been demonstrated as the genes for molecules which are drug targeting in cells (4).

It has been shown that the result of clinical PGx studies could be directly available for use in clinics. The consequences of PGx research affect the physician's prescription and

pharmacies policies for selling drugs and medicines besides avoiding the disadvantages of not efficient drug therapies in individuals, especially in cancer patients alongside the patients with common life style diseases as well as diabetes and cardiovascular disorders (5, 6).

Advanced technologies for PGx genotyping and data management

The emerging of high throughput genotyping technologies produced a huge amount of new findings for every part of the genetics area, including the pharmacogenomics and that reminds the necessity of the development of more comprehensive clinical guidelines in the field (7, 8). Alongside genome wide association studies (GWAS), investigations in a specific ethnic group seems beneficial for providing such guidelines too. Since, most of the pharmacovariants (genomic variants in drug-related genes) are introduced as rare (minor allele frequencies (MAF) <= 0.01) or extremely rare, sequencing based approaches considered as the better choice for uncovering such a population dependant variants (9). Specially, more comprehensive approaches like whole genome sequencing (WGS), whole exome sequencing (WES), and the utilization of long read sequencing are recommended and employed in recent years (10). However, functional assessment of not-interpreted pharmacovariants within the result needs the developing of more PGx dedicated bioinformatics tools, experimental methods, and artificial intelligence (A.I.) algorithms which are able to provide the functional prediction for every single pharmacovariant in a large-scale dataset (11-13). Furthermore, not only variants but also haplotypes play a significant role in enzyme activity and consequence phenotypes. Hence, the diagnostic technologies must take all

the functional variants within the pharmacogenes into account for allele imputation and diplotype identification and phenotype prediction (Figure 1) (14).

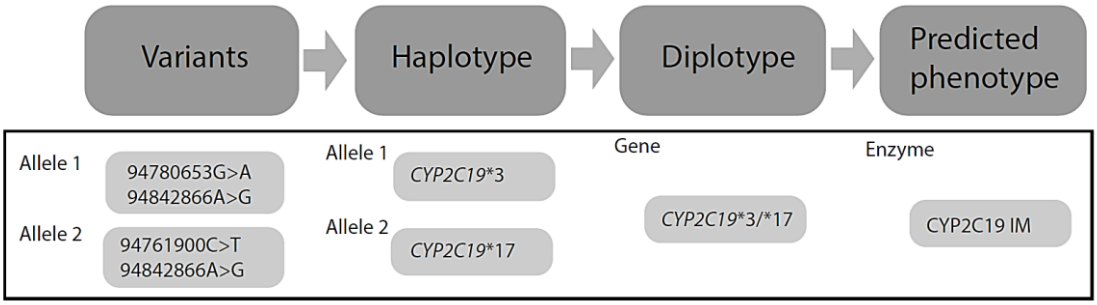

Figure 1: Pharmacogenomic genotype and phenotype assignment using the *CYP2C19* gene as an example: Variants per allele are assigned a haplotype according to the star(*)-allele nomenclature. The two haplotypes together form the diplotype which is subsequently translated into a predicted phenotype. adapted from van der Lee, M., Towards solving the missing heritability in pharmacogenomics, doctoral dissertation, Department of Clinical Pharmacy and Toxicology of Leiden University Medical Center. 2022: Leiden, The Netherlands.. *IM: intermediate metabolizer*.

Additionally, using eCARD or disease specific cards, displaying medical and health information has been developed in recent years as well. But so far, all these cards demonstrated the routine medical reports of the patients, such as blood pressure readings, immunization history, blood sugar, and cholesterol findings, surgery history, history of admission in hospitals, etc. Also, just a few set of pharmacogenes with actionable variants are already included in such cards (figure 2) (15, 16). But none of them specifically stored or mentioned the complete set of PGx data in an individual, containing both actionable pharmacovariants plus novel variants and/or not PGx interpreted markers, potentially influencial on drug response with the advantage of no necessity for updating the contains.

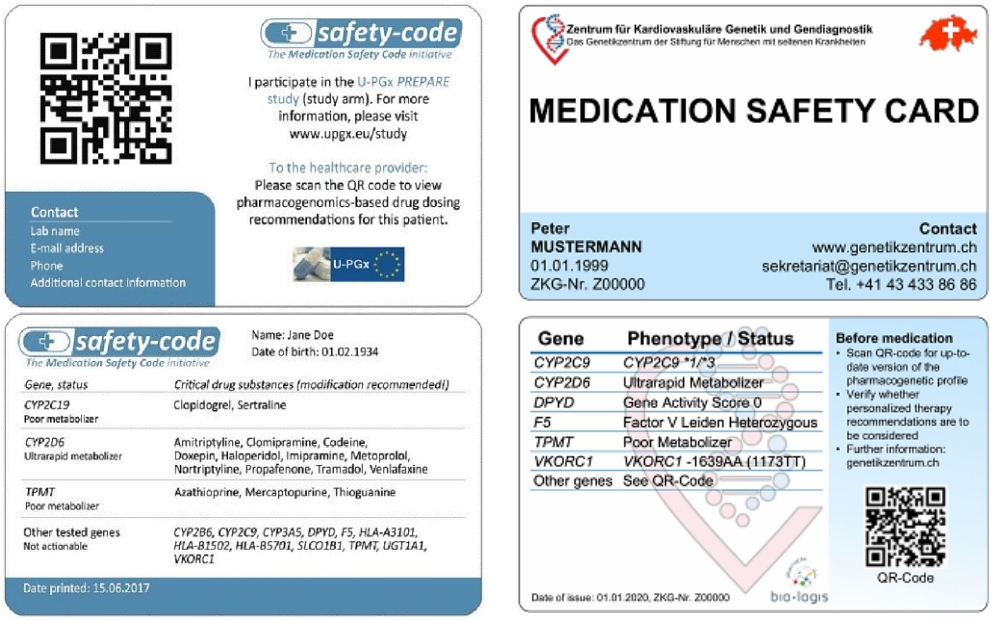

Figure2: Examples of previously developed PGx cards. Personalized PGx card with actionable variants and malfunction phenotypes for individuals. The QR code on cards will refer the clinicians to the particular webpage, containing the data for perscription modifications based on PGx reference databases guideline and recommendations. *Source: doi: 10.1371/journal.pone.0268534, doi: 10.3390/ijms21072308*
PGx: Pharmacogenomics

4. Study Aims:

Today, PGx introduced as the forefront of personalized medicine and proved it's important role for identification of functional genomic variants in drug-related genes. The first aim of this thesis, was to estabilsh a comprehensive map of the PGx data for a selected group of polish subjects, which directly affects the management and decisions of healthcare providers toward the personalized therapy in the clinic and pharmacies.

As a second aim, all the possible new genetic functional variations for the drug-related genes in the selected group of polish ethnic population would be identified and reported.

Finally, it would be possible for the study subjects to obtain portable medical records even when they are abroad and far away from their homeland, and provide them secure access to emergency PGx information, which could be considered as a wide-ranging therapeutic related data accessible via a custom designed ID card, carried in their wallet. This also could be seen as the future of personalized medicine implementation method through societies.

5. Materials and methods:

Biological samples for 100 participants collected from a long term biomedical study on local people named Bialystok plus. The consent forms were already handed to the subjects and the investigation confirmed by bioethical committee in Medical University of Bialystok, Poland (approval code: R-I-002/630/2018). DNA from cardiovascular patients and healthy people was extracted and genetic screening performed by WES platform in Novaseq6000 illumina instrument. Bioinformatic tools for primary data analysis utilized and the VCF files obtained respectively. The raw VCF files used in PGx-dedicated bioinformatic software and the result finalized by applying allele imputation approaches through reference databases including: PharmGKB, PharmVAR, CPIC, and DPWG. However, based on the tools documentation and instructions, some tools like Aldy and PharmCAT needed pre-processed data before running the software. The detail information on such tools are provided through the main text for the third article in this dissertation. On the other side, WES also provided the opportunity to extract the desired genes from raw VCF files in order to focus on particular set of genes, including those were related to patients' prescribed drugs, core pharmacogenes with already included guidelines and recommendations, and all drug-related genes in human obtained mostly from PharmGKB (Figure 3).

VCF files filtered by designed BED file for 1800 drug related genes and used for functional prediction by common bioinformatic tools as well as Sift, PolyPhen2, FATHMM, Provean, CAAD, Mutation Taster, etc. Also machine learning algorithm (Random Forest) employed for finding any hidden meaningful connection between genotype and patients phenotype through filterd VCF files.

Figure 3: Categorization of Pharmacogenes in WES data. Available genomic data for PGx analysis after running WES was including the genes with guidelines and recommendations, genes for target drugs in patients and all drug-related genes in human.
WES: whole exome sequencing. PGx: Pharmacogenomics

The latter is used for preparation of a manuscript, which is not included in current dissertation. Figure 4 and 5 illustrate the entire workflow for the study in this section with employed computational tools in each step.

Also, for PGx-dedicated bioinformatic algorithms specific data for nine core pharmacogenes with clear guidelines extracted and compared to explore the concordance and dicrepencies as well. Allele imputation performed through deep evaluation of the result and new insights for utilization of such tools for routine clinical PGx test reported (Figure 6 and 7).

In order to have a better view of the frequency, distribution, and potential function of identified pharmacovariants within less-studied drug-related genes or novel variants in core pharmacogenes, public freely available VCF files from various databases as well as Complete Genomics, Genome in Bottle, 1K genomes, ExAC, gnomAD, GET-RM, KAVIAR, etc. collected and target variants compared with annotated data from such sources. Variants with highly damaging scores undergone protein modelling for in-silico functional analysis and further evaluation of protein changes too.

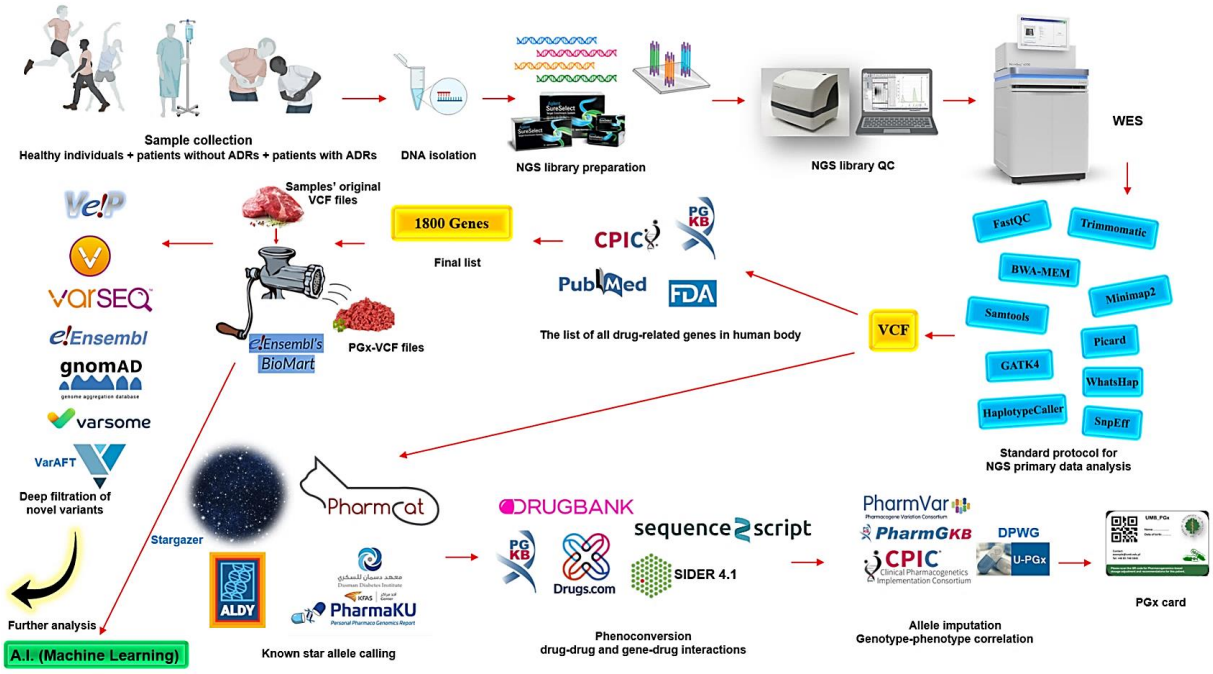

Figure 4: General view of workflow for current dissertation: genomic data obtained for PGx analysis after running WES for both patients and healthy individuals. VCF files used for PGx-dedicated bioinformatic tools directly. Also, filtered VCF files for 1800 drug-related genes used for data curation through deep computational and machine learning approaches. All the actionable and novel pharmacovariants and markers in drug-related genes stored in particular database and accessed by PGx card (see the text and third article for more details) . *ADRs: adverse drug reactions. WES: whole exome sequencing. PGx: Pharmacogenomics*

In case of no annotation for selected variants, neighbor markers and polymorphisms in related genomic coordinates assessed by KEGG and STRINGdb and ClinVar for cellular pathway and pothogenicity evaluations and estimation (gene walking approach). On this hypothesis, very close variants expected to show similar effects on protein structure and function. Figure 8 and 9 show these steps in details.

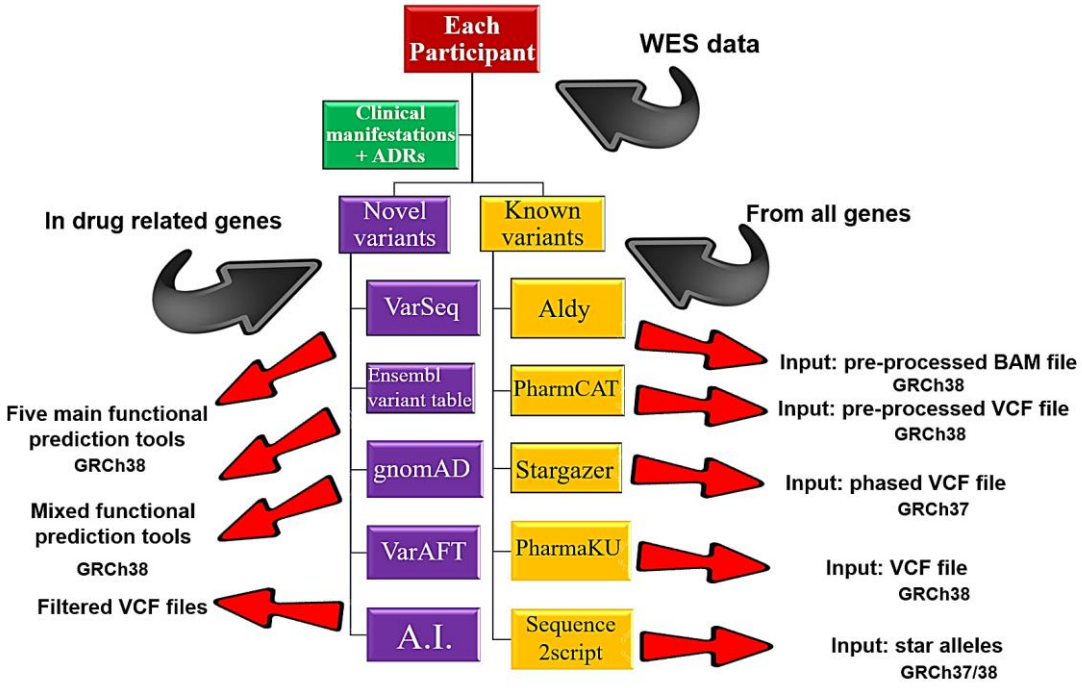

Figure 5: Approaches to deal with known and unknown genomic variants in drug-related genes: input data and various platforms, used in PGx data analysis for different types of variant within current study. After the registration of each participants demographic and clinical features (in case of patients), PGx-dedicated tools employed for known variants and common bioinformatic software and machine learning technologies for novel/not-annotated variants. (see the text and third article for more details) .
WES: whole exome sequencing. PGx: Pharmacogenomics

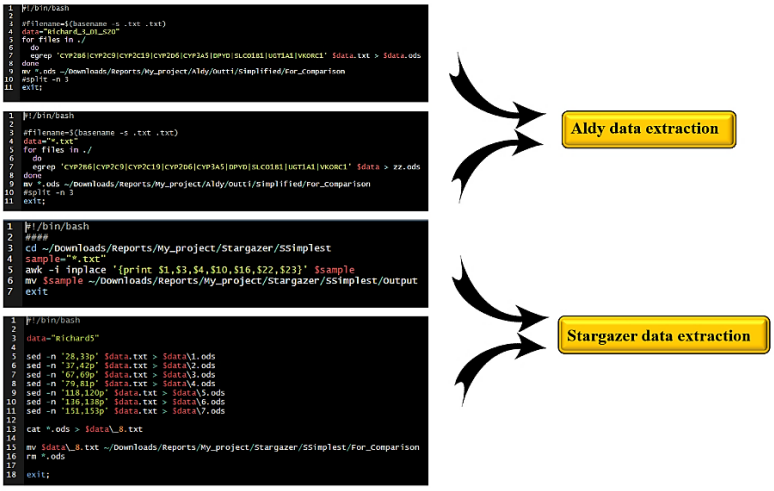

Figure 6: Data extraction from PGx-dedicated bioinformatic tools: Specific data for nine core pharmacogenes with guidelines and recommendations extracted and a comparison table made out of that for functional analysis of PGx-dedicated bioinformatic tools. *PGx: Pharmacogenomics*

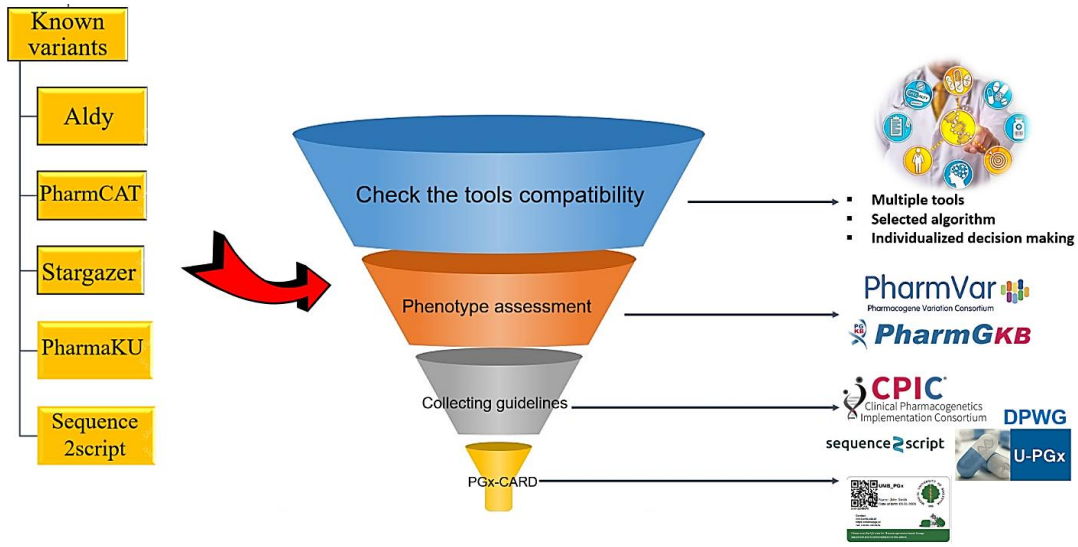

Figure 7: Known PGx variant analysis steps: After comparing and accuracy assessment of the result of PGx-dedicated bioinformatic tools for 100 WES data, phenotype prediction implemented through utilization of PharmGKB and PharmVAR databases. The related guidelines collected based on sequence2script and CPIC and DPWG recommendations and PGx card developed for participants. *PGx: Pharmacogenomics. WES: whole exome sequencing.*
CPIC: clinical pharmacogenetics implementation consortium
DPWG: dutch pharmacogenomics working group

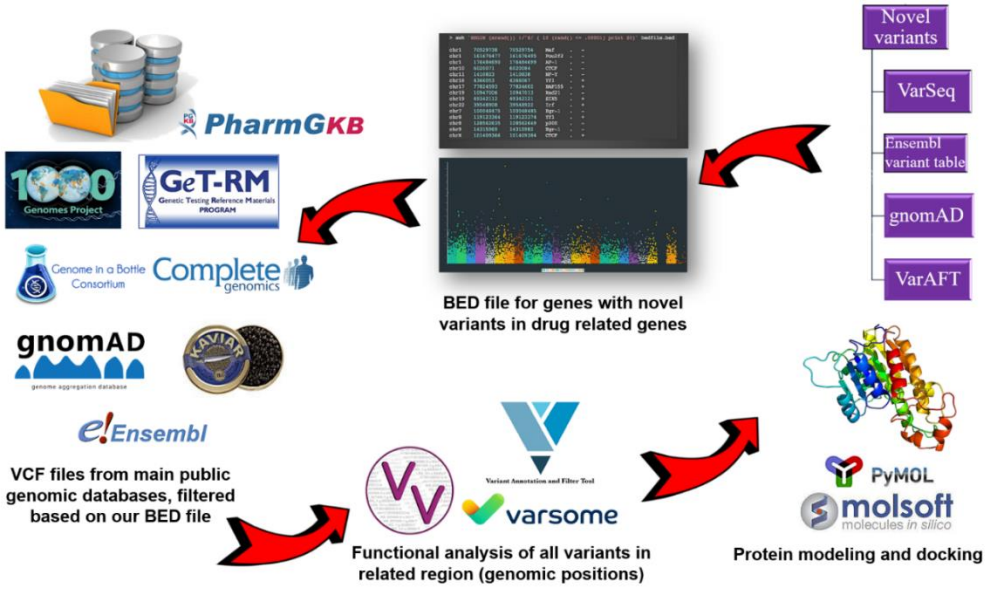

Figure 8: Novel/not-annotated pharmacovariant analysis steps (1): Initial steps included the utilization of specific BED file for selected identified variants and extraction of related genes and variants from public freely available datbases in order to have a general view on the frequency, distribution, and potential function of identified pharmacovariants in novel or not-annotated genes. *PGx: Pharmacogenomics*

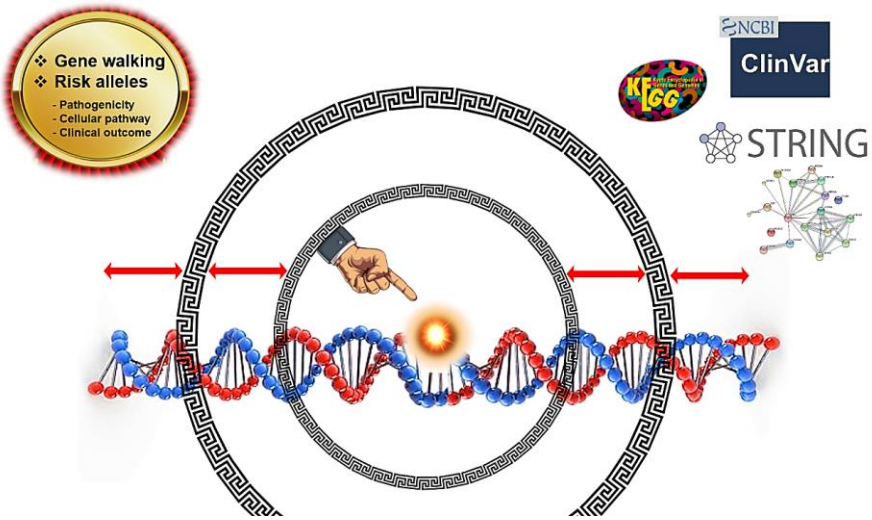

Figure 9: Novel/not-annotated pharmacovariant analysis steps (2): Variants with no PGx interpretation took the advantage of neighbor markers and polymorphisms with already contained annotation. If the data for closest variant was there then possible similar function may predict, specially when both variants were located in the same genomic coordination. The process called gene walking, which eventually lead to extraction of data for pathogenicity and cellular pathway for selected variants from main databases like KEGG, STRINGdb, and ClinVAR. *PGx: Pharmacogenomics*

6. Results and discussion of the articles included in dissertation :

Current dissertation provided the information on clinical PGx tests and Pharmacovariants in PGx screenings. State of the art advanced technologies for pharmacogenetic screening of patients (reactive test) and/or healthy individuals (pre-emptive genotyping) discussed too. Also, clinical PGx test implementation and data management through reference sources addressed by up to date approaches. While new PGx-dedicated bioinformatic tools and algorithms for data analysis scrutenized in detail, approaches to functional characterization and dealing with not annotated result in drug related genes introduced as well. Finally, digitalizing the result for immediate acess and utilization by healthcare systems evaluated and performed via creating local database. Main result of studies included in this dissertation are about PGx-related bioinformatic software outcomes and the approaches to deal with less-studied pharmacogenes and pharmacovariants.

Four publicly available PGx bioinformatics algorithms to assign PGx haplotypes were applied to nine selected very important pharmacogenes (VIP) and revealed a 45–70% concordance rate. To ensure availability of the results at point-of-care, actionable variants were stored in a web-hosted database and PGx-cards were developed for quick access and handed to the study subjects. Also, methods for deep computational filtration of large scale clinical PGx profiling introduced in order to perform functional assessment of not-interpreted pharmacovariants.

Numerous rare genetic variants within drug-related genes, anticipated to show important roles in variability in drug responses between individuals. Uncovering of such genetic markers are continuously increasing via the utilization of NGS technologies in clinic

(17). Advanced technologies for data mining and computational functional genomics paved the way of understanding the relationship between the genome and drug-related phenotype. Capability of computational approaches in drug repurposing for some particular medications demonstrated before (18). Current study confirmed that the utilization of multi bioinformatics tools and artificial intelligence approaches can help the discovery of novel rare pharmacovariants in NGS data and providing the link between genetic landscape and clinical manifestations for both rare and common variants within drug-related genes. However, clinical value and the utility of such approaches must be evaluated before heading toward the implementation through healthcare systems.

Indeed In-silico tools proved to be beneficial for big genomic data mining and addressing the detection of functional similarities between different genes and variants, categorizing and evaluation potential pathogenicity of novel/not interpreted variants, and also functional analysis of incidental findings (IFs) and variants with unknown significance (VUS) through various population with highest genetic diversity (19). However, not all of genomic variants (specially pharmacovariants) are placed within evolutionary conserved genomic locations and wouldn't be straight data for bioinformatic analysis. Hence, an adapted methodology for PGx data pre-filtration performed and employed several computational algorithms including novel approach of gene walking in order to focusing on identifiable genomic markers and their functionality assessments through extremely rare variants in pharmacogenes.

Registration of patients' genomic and actionable pharmacogenomic data in local electronic health record (EHR) has already implemented and novel digitalization

systems have been utilized for quick access to such data (20). Current study, also added detailed information on novel and/or not previously interpreted variants in less studied drug-related genes into a newly designed local database for participants' PGx actionable data. The information is including: applied genotyping technology, employed bioinformatics tools, variant genomic position, frequency, classification for pathogenicity, consequence, genome-built assembly, and functional assessment output based on both ACMG/ACGS guidelines. Also, considerations (for research use only) incorporated for possible interactions and conflicting with treatment outcomes plus links to update data on related gene in PhrmGKB.

However, while the comprehensive DNA sequencing approaches as well as WES and WGS can lead to more in-depth exploring of genomic and PGx data, some intrinsic challenges like VUS and IFs still pose problems through the result. Current study also brought several completely novel variants with no available primary annotation at all. The further process of such variants was not possible as our approaches were initially relying on already existed information for partially assessment at the beginning.

The utilization of advanced genotyping technologies in clinical PGx testing will need background knowledge and updated information on selection of appropriate instrument(s) and reference databases for data curation (21). The first article in the current dissertation will introduce different platforms of next generation sequencing (NGS) methods as the advanced genetic screening technologies and PGx related databases and dedicated bioinformatic tools for aquired data interpretation and functional characterizations. The pros and cons for each technique are displayed and advantages and current challenges for employment of such approaches is discussed.

Moreover, limitations of each NGS platform are discussed and solutions for managing and setting up the methods for clinical practice are addressed. The second article provides a guide on dealing with various types of variant, which investigators encounter while using NGS for PGx tests. Variant categorization and data sorting based on refrence guidelines explained and considerations by the test centers for functional assessments and diplotype/phenotype prediction are introduced. With this background, the third article demonstrates the real world PGx data out of WES on 100 Polish subjects. While the integration of NGS-guided drug stratification into daily clinical setting needs more in-depth studies on clinical validity and utility of already developed and available PGx dedicated bioinformatics tools for the implementation of clinical tests, this study adds important insights for utilization of such software through clinical practice and provides an algorithmic workflow for using on daily basis. How the tools work, what are the discrepancies, where do they come from, and what are the concordance rate for these tools are investigated and explained in details. Also, approaches to deal with novel and/or not interpreted variants in less studied drug-related genes within large scale genomic data introduced.

7. Conclusion:

- It is feasible to use clinical WES data for comprehensive PGx analysis and pharmacovariant detection.
- The discrepancies between state of the art PGx dedicated bioinformatic tools result may be managed through data comparison and curation, based on reference databases.
- Utilization of PGx-related bioinformatics algorithms plus PGx-oriented deep filteration of high throughput DNA sequencing result can help and accelerate the integration of PGx tests into daily clinical practice.

8. Summary of the results:

- Four publicly available PGx bioinformatics algorithms to assign PGx haplotypes were applied to nine selected very important pharmacogenes (VIP) and revealed a 45–70% concordance rate. Best tools selected for diplotype detection and phenotype prediction. PGx-cards were developed for quick access and handed to the study subjects.
- Deep computational filtration of large scale clinical PGx profiling through the utilization of BED file for (only) drug-related genes and applying multi-bioinformatic tools will help the performing of functional assessment of not/less-interpreted pharmacovariants.

9. Articles included in the dissertation

Review article

Applying Next-Generation Sequencing Platforms for Pharmacogenomic Testing in Clinical Practice

Alireza Tafazoli, Henk-Jan Guchelaar, Wojciech Miltyk, Adam J. Kretowski, Jesse J Swen*

Frontiers in Pharmacology – Pharmacogenetics and Pharmacogenomics section

DOI: 10.3389/fphar.2021.693453

IF: 5.988
MEiN: 100

frontiers in Pharmacology

REVIEW
published: 25 August 2021
doi: 10.3389/fphar.2021.693453

Applying Next-Generation Sequencing Platforms for Pharmacogenomic Testing in Clinical Practice

Alireza Tafazoli[1,2], *Henk-Jan Guchelaar*[3,4], *Wojciech Miltyk*[1], *Adam J. Kretowski*[2,5] *and Jesse J. Swen*[3,4]*

[1]Department of Analysis and Bioanalysis of Medicines, Faculty of Pharmacy with the Division of Laboratory Medicine, Medical University of Bialystok, Bialystok, Poland, [2]Clinical Research Centre, Medical University of Bialystok, Bialystok, Poland, [3]Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, Netherlands, [4]Leiden Network of Personalized Therapeutics, Leiden, Netherlands, [5]Department of Endocrinology, Diabetology and Internal Medicine, Medical University of Bialystok, Bialystok, Poland

Pharmacogenomics (PGx) studies the use of genetic data to optimize drug therapy. Numerous clinical centers have commenced implementing pharmacogenetic tests in clinical routines. Next-generation sequencing (NGS) technologies are emerging as a more comprehensive and time- and cost-effective approach in PGx. This review presents the main considerations for applying NGS in guiding drug treatment in clinical practice. It discusses both the advantages and the challenges of implementing NGS-based tests in PGx. Moreover, the limitations of each NGS platform are revealed, and the solutions for setting up and management of these technologies in clinical practice are addressed.

Keywords: pharmacogenomics, clinical implementation, next generation sequencing, clinical practice, PGx testing

OPEN ACCESS

Edited by:
Martin A. Kennedy,
University of Otago, New Zealand

Reviewed by:
Eng Wee Chua,
Universiti Kebangsaan, Malaysia
Chiara Fabbri,
King's College London,
United Kingdom

***Correspondence:**
Jesse J. Swen
j.j.swen@lumc.nl

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

Received: *11 April 2021*
Accepted: *26 July 2021*
Published: *25 August 2021*

Citation:
Tafazoli A, Guchelaar H-J, Miltyk W, Kretowski AJ and Swen JJ (2021) Applying Next-Generation Sequencing Platforms for Pharmacogenomic Testing in Clinical Practice. Front. Pharmacol. 12:693453. doi: 10.3389/fphar.2021.693453

INTRODUCTION

The Importance of Pharmacogenomics in Modern Medicine

Pharmacogenomics (PGx) utilizes individuals' genomic profiles to identify those who are at greater risk for adverse drug reactions or ineffectiveness. Many studies clearly indicate that drug-related genes, also referred to as "pharmacogenes," in the human genome contain extensive functional genetic variations (FGVs) and that different alleles are associated with diverse outcomes of drug treatments (Madian et al., 2012; Guchelaar, 2018; Suarez-Kurtz and Parra, 2018). Around 97–98% of people have at least one actionable FGV in their drug-related genes. In addition, the possibility of the presence of a genetic variant which could result in a loss of function (LOF) variant in pharmacogenes is 93% for every individual (Schärfe et al., 2017). Hence, the identification of the different genetic variants associated with the drug metabolism would impact on the prescription of medication, allowing for the selection of the right drug and dose, thereby reducing the potential adverse effects or

Abbreviations: ACMG, American College of Medical Genetics and Genomics; CADD, Combined Annotation-Dependent Depletion; CAP, College of American Pathologists; CLIA, Clinical Laboratory Improvement Amendments; CNV, Copy Number Variation; CYP, cytochrome P450; ExAC, Exome Aggregation Consortium; FDA, US Food and Drug Administration; GWAS, Genome-Wide Association Study; InDel, Insertion/Deletion; MAF, minor allele frequency; PROVEAN, Protein Variation Effect Analyzer; REVEL, Rare Exome Variant Ensemble Learner; SNP, Single Nucleotide Polymorphism; SNV, Single Nucleotide Variation; SV, structural variant; VCF, variant calling format; VEP, Variant Effect Predictor; VIP, Very Important Pharmacogenes.

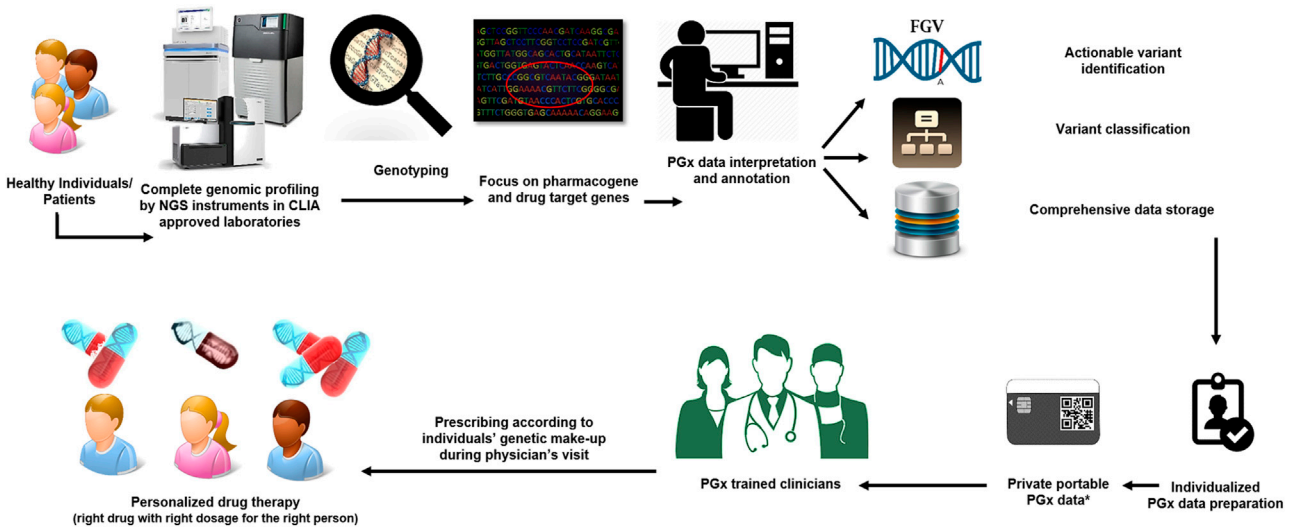

FIGURE 1 | A prospective view of the use of pharmacogenomics in modern medicine. Every person (sick or healthy individuals) will undergo comprehensive genomic screening before going to the physician/clinicians. The genetic variations in all pharmacogenes of an individual will be identified through data annotation and visualization by specific bioinformatics tools. The final report for each individual will be available through a private portable PGx electronic card. PGx trained clinicians will use an individual's genetic make-up report to tailor treatment to the patient's needs.

the therapeutic inefficacy. For clinical interpretation of PGx tests, the Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Dutch Pharmacogenetics Working Group (DPWG) guidelines are available as well as FDA drug-gene interaction recommendations. CPIC originally started as a shared project between PharmGKB and the Pharmacogenomics Research Network (PGRN) in 2009, and DPWG was launched in 2005 by the Royal Dutch Pharmacists Association. The two consortia have developed and published recommendations for numerous gene-drug interactions (JJ Swen et al., 2011; Relling and Klein, 2011). Both CPIC and DPWG provide updated, evidence-based, free access guidelines to facilitate and accelerate the establishment of a link between the results of PGx tests and specific dose recommendations. Nowadays, an increasing number of specified PGx tests are available in specialized CAP/CLIA approved clinical pharmacology/genome analysis centers around the world and can be found in the genetic testing registry (GTR, https://www.ncbi.nlm.nih.gov/gtr/) (Jiang and You, 2015).

The introduction of next-generation genome sequencing in PGx practice is an interesting and promising, albeit challenging, step. Currently, the field of PGx is moving from reactive testing of a single gene towards scanning an entire panel of genes involved in drug absorption, distribution, metabolism, and excretion (ADME) before prescribing (pre-emptive genotyping) by applying different types of next-generation sequencing (NGS) platforms (Bielinski et al., 2014; van Der Wouden et al., 2019). The results include all the PGx-related genetic variants in the genome which will be utilized to prepare drug dosing recommendations based on the predicted phenotype provided by the sequencing tests (**Figure 1**).

While the topic is highly popular and an overview of the current state of the NGS technologies for use in PGx testing has been offered in the literature previously (Schwarz et al., 2019), this article will discuss the challenges of detecting specific types of variants in PGx and interpreting such data in clinical practice. Solutions for the establishment and management of NGS devices in clinical practice are also addressed. A number of useful tables that provide detailed NGS-PGx-related information are also included. To aid with the terminology used throughout this manuscript, we included a concise glossary of NGS-related terminology in **Appendix 1**.

HOW CAN WE USE NGS FOR PGX ANALYSIS?

In this section, we firstly discuss the SNP-based PGx testing, which is currently the most frequently used test in the clinical PGx profiling of individuals, followed by targeted sequencing and whole-exome and whole-genome sequencing (WES/WGS).

SNP-Based PGx Testing in Clinical Practice

Fast, accurate, and inexpensive genotyping methods are key to the implementation of PGx in clinical practice. Currently, specific genotyping methods which mostly utilize different types of SNP-based genotyping approaches including real-time PCR with TaqMan probes and restriction fragment length polymorphism (RFLP) technique as well as gene panel-based genotyping methods such as ADME arrays are used in everyday clinical practice (Dorado et al., 2005; Johnson et al., 2012; Larsen and Rasmussen, 2017; Rasmussen-Torvik et al., 2017; Lemieux Perreault et al., 2018; Hippman and Nislow, 2019). In principle, genome-wide genotyping arrays such as Infinium Global Screening Array (GSA) could be used for routine PGx testing but are not yet commonly applied for

TABLE 1 | Summary of the recent studies that used the NGS technologies for functional PGx variant detection.

Study objective	n =	Applied NGS platform(s)	Covered drug-related genes	Identified variants	Result	Reference
Platform validation and variant discovery	3 × 96	Targeted sequencing	84	SNVs	A custom-designed panel (PGRNseq) could be an ideal platform for both the common and the rare PGx variants identification in large cohorts and suitable for the clinical tests	Gordon et al., 2016
Platform validation and variant discovery	376	Targeted sequencing	114	SNVs	Targeted sequencing panels are ready-to-use platforms for comprehensive pharmacogene profiling including common plus rare variants in ADME core genes towards the implementation of the personalized medicine	Han et al., 2017
Platform validation	2 (cell culture)	Targeted sequencing	3	SNVs CNVs InDels	Variants and haplotype detection of challenging ADME genes were successfully achieved	Ammar et al., 2015
Platform validation and variant discovery	235	Targeted sequencing	100	SNVs CNVs	Designed PGxSeq panel with high accuracy identified clinically relevant variants in 39 genes including CYP2D6 CNV and UGT1A1*28 TAA repeats in the promoter. The allele frequency and the homozygosity were also determined	Gulilat et al., 2019
Platform validation and variant discovery	150	Targeted sequencing	340	SNVs Small InDels	Panel-based NGS pipeline developed and revealed 7,273 novel variants in 340 ADME genes of 150 Caucasian liver donors with an accuracy of >99%. The functional prediction allowed for the prioritization of the variants for further analysis	Klein et al., 2019
Validation of known variants	60	Targeted sequencing	20	SNVs InDels	Prediction model of the atorvastatin plasmatic concentrations in healthy volunteers through the sequencing results explained well	Cruz-Correa et al., 2017
Platform validation	98	Targeted sequencing and WGS data	19	SNVs CNVs	The concordance between the two platforms estimated to >97% for identified variants. The CNVs concordance in CYP2D6 gene also demonstrated 90% of accuracy. 95 children had at least one clinically actionable pharmacogenetic variant	Cohn et al., 2017
Validation of known variants	1,583	Whole-exome sequencing data	11	SNVs	At least one actionable phenotype was present in 86% of individuals. Repurposing WES data can yield meaningful pharmacogenetic profiles for 7 of 11 important pharmacogenes, which can be used to guide the drug treatment	van der Lee et al., 2020a
Validation of known variants	94	Whole-exome sequencing	3	SNVs	Diagnostic genotyping identified PGx variants in CYP2C19, CYP2C9, and VKORC1 genes in 91% of all cases. Of this, 20% indicated potential immediate effects on the currently used medications	Cousin et al., 2017
Platform validation	36 + 12	Whole-exome sequencing	36	SNVs InDels	High concordance revealed through cross-comparison of WES and other platforms as well as the MiSeq amplicon sequencing data and the IPLEX ADME PGx panel. WES was introduced as a promising tool in PGx profiling with a low error rate of <1%	Wee Chua et al., 2016
Platform and discovery rate validation	2504 of WGS data + 59,898 of WES data	WGS and WES data	208	SNVs CNVs	The population-specific deletion and the duplications were revealed in 97% of the analyzed subjects and the related frequencies were reported and confirmed via Sanger sequencing	Santos et al., 2018
Platform validation and variant discovery	1,000 Genomes data	Whole-genome sequencing data	160	SNVs InDels	Putatively functional variants within known pharmacogenomics loci identified that could account for association signals and represent the missing causative variants underlying drug response phenotypes	Choi et al. (2019)
Variants validation and discovery	547 individuals from in-house cohort data + gnomAD data	Whole-genome sequencing data	11	SNVs InDels	For improved precision medicine, PGx testing should move towards WGS-based	Caspar et al. (2020)

(Continued on following page)

TABLE 1 | (*Continued*) Summary of the recent studies that used the NGS technologies for functional PGx variant detection.

Study objective	n =	Applied NGS platform(s)	Covered drug-related genes	Identified variants	Result	Reference
					approaches as a feasible and most comprehensive method	
Platform validation	44,000 biobank participants	WGS and WES data + microarray data	11	SNVs CNVs	WGS and microarray demonstrate more concordances for the obtained results. WES is not suitable for PGx preemptive predictions. However, the microarrays are more cost-effective than the sequencing platforms. Overall, the implementation of the PGx tests and the recommendations may affect at least 50 daily drug doses per 1,000 inhabitants	Reisberg et al. (2019)
Variant discovery	3	Targeted sequencing	16	SNVs	The functional alterations and variants with potential impact on anti-TNF drug response successfully introduced by rapid, sensitive, and cost-effective NGS-based pharmacogenetics methodology	Walczak et al., 2019
Variant discovery	392	Whole-exome sequencing	21,000	SNVs InDels	Exome sequencing revealed novel genetic loci with a strong association with on-treatment reactivity and hereditability of platelet and clopidogrel response	Price et al., 2012
Variant discovery	482 + 7	Whole-genome sequencing	231	SNVs InDels Tandem substitutions	17,733 ADME variants/individuals detected. In addition to known PGx markers, 1,012 novel variants with potential deleterious function identified in exons, introns, gene promoters, and proximal regulatory regions	Mizzi et al., 2014
Variant discovery	100	Whole-genome sequencing	437	SNVs	The analysis revealed 227 common and 466 rare population-specific potentially functional SNVs	Sivadas et al., 2017

this purpose. While the technology is still developing, the main limitation is that the identification of the structural PGx variants such as Copy Number Variations (CNVs) and hybrid genes as well as CYP2D6/7 is mostly ignored. Moreover, the variants in the pharmacogenes that are tested are limited to currently known and common alleles. Although several versions of arrays are being enriched with more specific PGx variants (thousands of drug-related biomarkers) (Arbitrio et al., 2016; Thermofisher.com/Pharmacoscan, 2018; Illumina, 2020), no phasing information will be obtained through these tests, which makes it more challenging to provide an accurate phenotype prediction.

Hence, the properties of NGS technologies make them an interesting approach to performing clinical PGx testing. In recent years, several investigators have explored different approaches utilizing NGS platforms, namely, targeted sequencing, WES, and WGS in pharmacogenomics. **Table 1** shows several studies stratified by different approaches.

Targeted Sequencing Panels

Research into PGx over the years has resulted in the identification of numerous genes which may play an essential role in drug metabolism, transport, and targeting in the human body. However, not all of them are strongly associated with drug response phenotypes and therefore CPIC and DPWG only provide clinical recommendations for specific variants in well-known pharmacogenes.

Gordon et al. developed the PGRNSeq panel as a balance between cost, throughput, and depth of coverage. The panel included clinically actionable CPIC genes as well as genes for which little was known, although a primary association with the PGx trait existed. It was concluded that the PGRNSeq panel is suitable for both the clinical investigations and the discovery studies. However, some non-coding parts and complex structural variants for specific pharmacogenes (including CYP2A6, CYP2D6, and HLA-B) alongside better computational resources for data interpretation remain to be developed. In a similar approach, Han et al. developed an unbiased and broad-range NGS panel and suggested that the utilization of such panels may be a valuable tool in the comprehensive study of PGx genes. The selection of genes for inclusion in the panel was based on the pharmaadme and Www.pharmaadme.org database (Gordon et al., 2016; Han et al., 2017).

Customized PGx panels can also serve as a highly accurate approach to variant detection in the clinical PGx testing. Gulilat et al. developed a targeted exome panel, named PGxSeq, for capturing both SNVs and CNVs in pharmacogenes. They demonstrated that PGxSeq could be employed as a reliable tool for common and novel SNVs alongside CNV detection in pharmacogenes in clinical use. However, a limitation of the work was that the validation was restricted to 39 loci in 16 genes in specific population samples. Moreover, pharmacogenetic variants in non-coding and regulatory parts were not included (Gulilat et al., 2019). A comprehensive PGx panel that includes all coding

regions, adjacent introns, and 5′ and 3′ UTRs in flanking sequences of 340 ADME genes has recently been developed by investigators in Germany. The identification of genes for inclusion in the panel was based on multiple sources including PharmaADME, PharmGKB, and ADME-related genes from the literature. Compared with other genotyping methods, accuracy was high, with >99% correct calls. The obtained data allowed for the covering of coding and functional non-coding parts and provided related data for both common and rare variants in addition to revealing novel associations. The detection of some limited InDels and integration of rare variants into PGx by the current computational predictors alongside the sample size were reported as limitations of the panel (Klein et al., 2019).

Long-Read Sequencing for Gene Panels

Several PGx genes involve complex variants such as tandem repeats, pseudogenes, and CNVs. Long-read sequencing approaches (on average over 10 kb in one single read) have been used previously in the profiling of different complex genomic loci and have been proposed for the identification of such challenging genomic areas in PGx (Ardui et al., 2017; Mantere et al., 2019; van der Lee et al., 2020a). In this field, Ammar et al. applied long-read sequencers to identify PGx variants and haplotypes in three challenging pharmacogenes: CYP2D6, HLA-A, and HLA-B. The constructed haplotypes were confirmed by HapMap data and statistically phased Complete Genomics (WGS data from the public 69 genomes project) and Sequenom genotypes (for 36 SNP, InDels, and CNVs for CYP2D6). The results demonstrated the potential of long-read sequencing in clinical PGx (Ammar et al., 2015). In addition to haplotyping, variant phasing is also a challenge in PGx. Long-read sequencing has also been employed to resolve phasing issues and provide a solution to the accurate genotyping of complex PGx genes. Yusmiati Liau et al. utilized the GridION platform for sequencing and haplotyping of the entire CYP2D6 gene. Known and new alleles and subvariants plus duplicated alleles were assigned accurately with correct phasing. The approach also demonstrated the capability of processing multiple samples simultaneously and appeared to be a time- and cost-effective method (Liau et al., 2019).

Whole-Exome Sequencing

More comprehensive methods such as WES and WGS identify high numbers of pharmacogenetic biomarkers. In addition, these sequencing approaches may facilitate the discovery of novel loci. While it is possible to reuse WES for PGx purposes for known variants, the application for novel variants is challenging as the investigators would need a confirmative study or extensive *in-vitro* research to attribute potential, newly identified variants in a particular gene to drug response. This is particularly true if it is not clear what functional effect the genetic variation exerts on protein function and/or expression. Van der Lee et al. investigated the feasibility of repurposing WES data for the extraction of a PGx panel of 42 variants in 11 pharmacogenes to provide a pharmacogenomic profile. Based on the Ubiquitous Pharmacogenomics (U-PGx; www.upgx.eu) panel which includes all the actionable genes and variants in the DPWG guidelines, the authors successfully extracted information regarding 39 variants out of the total 42. At least one actionable phenotype was present in 86% of the analyzed data from the included subjects. Although structural variants (SVs) and copy numbers in some pharmacogenes as well as CYP2C19, UGT1A1, CYP3A5, and CYP2D6 were not detected, and the study suffered from a small number of drug-related genes and a limited sample size, the authors concluded that the WES data can yield meaningful pharmacogenetic profiles for 7 out of 11 important pharmacogenes (van der Lee et al., 2020b). To assess the potential benefits and the limitations of using the clinical WES data for PGx analysis as a secondary finding, Cousin et al. analyzed the clinical WES data for the detection of any FGVs in three important pharmacogenes. PGx variants were extracted from the WES test results of patients and used in addition to their medical history data. A pharmacist interpreted the PGx data based on multiple resources including CPIC, UpToDate, Micromedex, and AskMayoExpert and used the information to perform a genotype-informed medication review. The authors concluded that PGx testing early in life would be helpful for prescribing physicians to make future prescribing decisions (Cousin et al., 2017). The accuracy and the concordance rate for the WES variant calling were also investigated by Wee Chua et al. The researchers performed a cross-comparison between the WES and MiSeq amplicon sequencing data in addition to the WES and iPLEX ADME PGx panel in 36 and 12 samples, respectively. The rate obtained for both comparisons was high (99%), which indicates that WES is a promising tool in PGx profiling of individuals with an estimated error rate of <1% (Chua et al., 2016). However, despite these positive results, an important limitation of WES is that several important PGx variants, including CYP2C19*17 and VKORC1, are located outside of the captured regions of routine whole-exome sequencing.

Whole-Genome Sequencing

Complete genomic variants (including PGx-related markers) for an individual would be available through the utilization of the WGS approach. Although the big data interpretation of such tests is still challenging, a decrease in sequencing costs alongside the comprehensiveness of WGS may result in the method becoming a standard platform for clinical PGx tests.

Through using the WGS data from phase 1 of the 1,000 Genomes project and subsequent annotation, 69,319 variants including SNVs (94%) and InDels (6%) were revealed in 160 pharmacogenes (127 CPIC genes and 64 VIP genes from PharmGKB). Minor allele frequency for the variants was >1%, of which 8,207 were in strong linkage disequilibrium (LD) (r^2 > 0.8) with known PGx variants. The alterations were distributed in various parts of the genome including intronic, coding, and 5′ upstream and 3′ downstream regions. In the end, the authors identified putatively functional variants within known pharmacogenomic loci underlying drug response phenotypes and suggested direct testing instead of relying on LD, which is going to be different among populations. A limited sample size and exclusion of rare variants (MAF <0.01) in addition to a lack of an experimental validation study were reported as the main

limitations of the investigation. However, the results from such PGx studies facilitate the translation of the findings of the genomic analysis into clinical practice (Choi et al., 2019). While the known PGx gene panels could be included in the WGS data and considered a source for clinical PGx and drug prescribing, the remainder of the information could still be useful for discovery studies.

The functional CNVs in ADME genes are distributed with significantly different frequencies across diverse populations (He et al., 2011; Martis et al., 2013). The NGS data could also be used for CNV calling in different ethnic backgrounds. The investigators used the integrated WGS and WES data from 1,000 Genome and ExAC repositories for CNV identification in 208 pharmacogenes. Novel CNVs (deletion in 84% and duplications in 91% of genes) across six different populations of non-Finnish Europeans, Africans, Finns, East Asians, South Asians, and admixed Americans were decoded successfully. The final result highlighted the necessity for the comprehensive NGS-based genotyping of the pharmacogenes for the CNV identification alongside their allele frequencies. The assessment of the contribution of such CNVs to the drug response outcomes is also possible through a population-specific analysis of rare variants (Santos et al., 2018). Applying NGS for recognizing the actionable variants in genomic profiles may lead to lifetime utilization of PGx information for related individuals. Furthermore, future bioinformatics tools could potentially be utilized for the NGS data re-analysis and the functional prediction of novel variants (Cousin et al., 2017).

As demonstrated, the targeted sequencing approaches are most suitable for genotyping of known PGx genes, including the low-frequency variants. For the discovery of novel pharmacogenes of interest, WGS and WES are considered better choices (Reisberg et al., 2019). WES and WGS also offer the possibility of data repurposing, which means that the clinicians can benefit from the existing clinical sequencing data to extract a PGx profile to inform drug treatment. Although the NGS data from different platforms offer many potential benefits, there are still several challenges and limitations which are discussed in the following sections.

CHALLENGES IN THE APPLICATION OF NGS PLATFORMS FOR THE DECODING OF PGX VARIANTS IN SPECIFIC PHARMACOGENES

From the studies presented above, it appears that most types of variants in the coding and non-coding or regulatory parts of drug-related genes including SNVs, InDels, CNVs, and some structural alterations such as tandem substitutions could be identified with NGS, particularly with long-read sequencers and WGS. However, some well-known clinically actionable pharmacogenetic variants still pose a challenge for the NGS methods. Challenging genes include some core ADME genes, such as CYP2D6 which contains many different known (>100 * alleles, www.pharmvar.org) variants in different populations. Moreover, high sequence similarity and genetic recombination between real genes and close pseudogenes, such as CYP2D7 and CYP2D8, structural rearrangement complexities, and high CNVs among individuals present substantial challenges. Here, the routine short-read NGS approaches will not clarify the genetic profile of an individual and offer proper phenotype prediction. Furthermore, difficulties in the alignment procedures make interpretation and translation into clinical use complicated. Although some of these problems can be resolved by high-resolution techniques, including long-read sequencing, such sequencers with lower error rates (as well as PacBio Sequel HiFi II) are only available through highly specialized centers and are not yet applied in routine clinical practice (Yang et al., 2017). In addition, the technology is currently not being considered for the large-scale genome analysis in the PGx studies (van der Lee et al., 2020a).

Another example of a challenging pharmacogene is UGT1A1, with some important variants in the non-coding parts of the gene (TA repeats in the promoter of the gene, particularly UGT1A1*28, which affect the gene transcription and hence enzyme activity) (Bosma et al., 1995; Dalén et al., 1998; Numanagić et al., 2015). The gene harbors more than 113 functionally relevant variants, most of which reduce or enhance enzyme function, in addition to many other variants with unknown significance. The allele frequency is heavily population-specific, too. However, most of the panels focus on commonly known genotypes and could easily miss predictive variants in particular cases. By way of illustration, FDA approved the test for *28 allele but not *6 allele for irinotecan, although the latter is the main cause of the altered activity of the UGT1A enzyme in the Asian populations (Ikediobi et al., 2009). Also, the utilization of more comprehensive platforms such as WES is accompanied by poor and insufficient coverage for non-coding parts, which may result in the lower concordance and weak diplotype and CNV calls for the UGT1A1 gene (van der Lee et al., 2020b).

A third challenging region is the HLA genes. They are characterized by high sequence homology and prone to error in the capturing procedure and possible misalignment in the mapping processes. In addition, more than 21,000 known alleles and several pseudogenes and some InDels in the intronic regions of HLA class I and class II genes require the utilization of a proper platform, and more advanced IT infrastructure for the bioinformatics analysis and the identification of various potential predictive PGx markers, particularly in the newly studied populations (Klasberg et al., 2019). HLA alleles are important not only in PGx but also in other medical fields, including the genomic evaluation of multifactorial disorders and organ transplantation. Unfortunately, most of the HLA variants are rare and population-specific and are not included in routine clinical PGx testing (Nakkam et al., 2018). Today, many bioinformatics tools and algorithms available for HLA variant calling and haplotype phasing based on the WGS, WES, and targeted sequencing results. However, the high coverage of the genomic region is preferred as input for the allelic imputation by most software (Karnes et al., 2017). The available tools and their pros and cons have been discussed

TABLE 2 | Pharmacogenes with the associated challenges that render them difficult to genotype.

Gene	Challenge(s)	Reference
CYP2D6	–Structural variants and gene rearrangements –Pseudogenes –Copy Number Variations –Presence of novel variants –Highly polymorphic region –Substrate-specific effects of some alleles	Taylor et al. (2020) PharmVar structural variations CYP2D6
UGT1A1	–Rare population-specific variants –Variants in non-coding parts of the gene –Independent haplotypes with less linkage disequilibrium	Barbarino et al. (2014) Marques and Ikediobi, 2010
VKORC1	–Important variants in non-coding parts of the gene	Saminathan et al. (2010) Owen et al., 2010
HLA	–Rare population-specific variants –Highly polymorphic regions	Illing et al. (2017) Klasberg et al., 2019
SLC6A4	–Rare population-specific variants	Lam (2013)

comprehensively in the literature (Ka et al., 2017; Kawaguchi et al., 2017; Xie et al., 2017). In general, to overcome the challenges of decoding PGx variants in specific genes, up-to-date knowledge of PGx-related genomics for physicians requesting the test in addition to the selection and utilization of an appropriate platform and interpretation tools for each situation by PGx test centers is required. This may also include previous knowledge of some particular PGx alleles with substrate-specific effects. For example, CYP2D6*17 encodes an enzyme with an increased capacity to metabolize haloperidol but an impaired ability to metabolize codeine (Oscarson et al., 1997; Wennerholm et al., 2002). In addition, occasional discrepancies between guidelines on the classification of genotypes into metabolic groups (which is key to formulating corresponding therapeutic recommendations) must also be considered (Caudle et al., 2020). **Table 2** summarizes some challenging pharmacogenes and their main features that need to be taken into consideration during sequencing or panel design.

CHALLENGES AND OPPORTUNITIES FOR DATA ACQUISITION AND INTERPRETATION

The NGS data annotation, in the form of PGx phenotype prediction, is a highly specialized task that requires both molecular knowledge and clinical knowledge. The extraction of actionable, putative, or likely pathogenic variants from large, sophisticated raw data requires considerable time and effort as well as accurate validation methods. The current approaches include newly developed PGx dedicated tools for star allele calling in pharmacogenes (discussed in the following sections). Here, we address the key considerations, discuss some features of the common PGx-related tools, and propose solutions for managing the challenges.

Targeted Sequencing Panels

Unlike with other genotyping approaches, performing a sequencing run always offers the possibility of decoding novel variants in the sequenced part(s). This has also been observed in the targeted sequencing panels of known pharmacogenes, where novel variations appeared in addition to common markers (Gulilat et al., 2019). Indeed, the variants with unknown clinical significance (VUS) in the NGS data and with no clear connection to pharmacogenetics present a real challenge as far as the implementation of such technologies in clinical practice is concerned. Nevertheless, handling VUS as potentially important identified variants is essential since if appropriate approaches to the correct interpretation were not available, the real functional alleles might simply be introduced as non-actionable. Therefore, a prediction is not feasible easily on the functionality of VUS to interpret the potential effects on the drug responses in a patient. However, because of the lower number of such findings in panels, replication and validation studies using other orthogonal genotyping methods, in silico algorithms, genetic screening for first degree relatives of the proband, and use of GWAS, HapMap, or gnomAD datasets for meta-analysis will be faster and more easier with regard to predicting and confirming the negative or neutral functionality of variants and demonstrating the phenotype associations in the targeted sequencing approaches (Svidnicki et al., 2020).

Whole-Exome and Whole-Genome Sequencing

As expected, VUS are more common in WES and WGS. The situation becomes even more complicated when the results involve novel PGx genes. Online tools such as SIFT and PolyPhen2 as well as other algorithms, including CADD and PROVEAN, plus Ensembl based sources with multiple integrated tools like VEP and REVEL, are available for the prediction of the damaging effects of a large number of variants. However, these tools rely primarily on evolutionary conservation and utilize amino acid or nucleotide sequence alignment, which is less applicable to pharmacogenes. Also, low predictive value of these tools has recently been demonstrated (Lee et al., 2019; Zhou et al., 2019).

Furthermore, incidental findings (IFs), referred to as secondary findings in the ACMG recommendations (Kalia et al., 2017), can be expected in different types of high

TABLE 3 | Key features of the PGx dedicated variant functional prediction tools.

Tool/Algorithm	Main features	Reference
Stargazer	Stargazer calls the star alleles from the NGS data by detecting SNVs, InDels, and structural variants. Stargazer detects variations with structural changes including gene duplications, deletions, and conversions by calculating the paralog-specific copy numbers from read depth	Lee et al. (2019)
PharmCAT	Pharmacogenomics Clinical Annotation Tool (PharmCAT) captures the variants indicated in guidelines from a genomic data set derived from sequencing or genotyping technologies (i.e., VCF), infers haplotypes and diplotypes, and generates a report containing genotype/diplotype-based annotations, as well as guidelines and recommendations according to CPIC guidelines	Sangkuhl et al. (2020)
Aldy	Aldy is a computational tool that performs allelic decomposition of highly polymorphic, multi-copy genes through the use of the whole or targeted genome sequencing data and identifies multiple rare and novel alleles for several important pharmacogenes	Numanagić et al. (2018)
Astrolabe	Astrolabe (former Constellation) is a computational method and probabilistic scoring system that enables automated ascertainment of CYP2D6 and CYP2D19 activity scores from the unphased NGS data, aligned with the catalog of pharmacogenetic alleles with high percentage of analytic sensitivity and specificity	Twist et al. (2017)
Cypripi	Cypripi is an algorithm that computationally assumes CYP2D6 genotype at base-pair resolution from the high throughput sequencing data. It can resolve complex genotypes, including the alleles that are the product of the duplication, deletion, and fusion events involving CYP2D6 and its related pseudogene, CYP2D7	Numanagić et al. (2015)
g-Nomic	g-Nomic is PGx interpretation software that provides recommendations on the suitability of a given combination of drugs for each patient according to their genes and polymedication	Sabater et al. (2019)
PHARMIP	PHARMIP uses drug modeled structure and up-to-date bioinformatics tools and/or databases to understand the genetic factors that cause drug-related adverse reactions	Zidan et al. (2020)
Cyrius	Superior, accurate genotyping of CYP2D6 compared to other existing methods as well as Aldy and Stargazer. All types of variants and haplotype calling in addition to the structural and homology analysis will be covered for both GRCh38 and 37 genome builds	Chen et al. (2021)

throughput sequencing and genotype screening methods. They are mostly defined as annotated functional variants in major drug-related genes which were not expected in the specified assessment but may be either related or unrelated to the particular medication taken by the patient. This adds to the complexity of reporting findings from PGx profiling, where the DNA variants may alter the drug efficacy or increase the risk of serious adverse drug reactions. Such findings could be reported as variants with potential usage in guiding therapy if they are managed properly through appropriate clinical genomic assays, vigorous genotype-phenotype correlation studies, and utilization of PGx-related sources for data interpretation and variant scoring (Lee et al., 2016). However, the existence of secondary findings would also be associated with some technical issues in the employed NGS platform. These issues include the percentage of coverage and type of sequencing methods as well as the number of evaluated individuals, evaluation of family members or randomly selected patients (Westbrook et al., 2013). Yet, not all secondary findings that are identified need to be reported in the result of a clinical test. The ACMG also declared a policy statement for reporting particular secondary findings in the clinical setting (L Blackburn et al., 2015; Miller et al., 2021). However, the statement is related to non-PGx secondary findings. Moreover, many pharmacogenetic variants are not disease-causing. Therefore, the relevance of reporting secondary findings may not be obvious at the time of submitting the report, particularly when only a specific set of pharmacogenes is tested. For the pharmacogenes connected with disease risk, the secondary findings may be handled in accordance with the current ACMG recommendations; that is, it is not necessary to provide a separate set of recommendations for those genes. Nevertheless, while the purpose of PGx testing is to exhaustively (and pre-emptively) profile genes that may potentially alter the drug response, curating and storing the information relevant to the future drug therapy may indicate that no findings should be considered "secondary," particularly when untargeted methods as well as WES and WGS are employed.

Recently Developed Bioinformatics Algorithms for PGx Variant Calling

Concentrated efforts have been undertaken to design and develop specific PGx tools for the identification of SNVs, CNVs, structural rearrangements, gene deletion, gene duplication/multiplication, haplotype phasing, diplotype calling, and phenotype prediction out of the NGS data in the clinical setting. The tools as well as Stargazer, PharmCAT, Astrolabe, Aldy, Cypripi, include special algorithms, which were designed for the interpretation of the PGx variants (Numanagić et al., 2015; Twist et al., 2016; Klein and Ritchie, 2018; Numanagić et al., 2018; Lee et al., 2019). Furthermore, some other tools including g-Nomic and PHARMIP were developed for providing recommendations based on the general information obtained from a PGx test (Sabater et al., 2019; Zidan et al., 2020). The advantages and the disadvantages of each of the tools have been demonstrated previously in the literature (Twesigomwe et al., 2020). **Table 3** provides a concise overview of the key features of these tools. Stargazer, Astrolabe, and Aldy have been fully analyzed and are widely used in the field. Twesigomwe and colleagues have recently performed a comprehensive and

systematic comparison of the functions of these three tools in calling different CYP2D6 variants. The results of the study demonstrate that Aldy and Astrolabe are better common and rare SNV callers compared to Stargazer. Yet, Stargazer outperformed the other tools in rare homozygous allele phasing due to its in-built supplementary algorithm. Calling InDel star alleles in the short-read NGS data and the hybrid rearrangements was challenging for all three algorithms. For other structural variants, gene deletion, duplication, and multiplications, Aldy demonstrated higher concordance in comparison to Stargazer and Astrolabe, respectively. Noticeably, Astrolabe performed weak structural variant calling in comparison to the other two tools. Although Stargazer displayed better performance in CNV calling and the identification of hybrid rearrangements, it simultaneously revealed the highest number of non-genotyped diplotypes for the samples including structural variants. Unfortunately, all three tools had difficulty calling diplotypes with high copy numbers. While these genotypes are very rare, they may still be considered an important variant in some isolated populations. The phenotype prediction and the clinical accuracy of Aldy, Astrolabe, and Stargazer were also evaluated. Remarkably, the concordances were higher than the diplotype concordances as the activity scoring systems may assign the same values as the true function of the wrongly genotyped samples. The impact of the sequencing coverage and the misalignment of InDels on genotyping accuracy was also investigated. The study, however, had some limitations. It used simulated data for most rare and structural variants, did not compare the performances of the three tools across the NGS data from the targeted custom-capture panels, and did not compare the impacts of different aligners on the variant calling processes. Novel SNVs calling was also not analyzed in the study and reliable validation studies were not included (Twesigomwe et al., 2020). Aldy and Stargazer may also result in false-positive/false-negative results in small variant calling, since they rely on initial read alignments. Another major obstacle is that two of the three tools does not support the GRCh38 genome assembly and that the investigators may need to lift their alignments to GRCh37 (i.e., https://genome.ucsc.edu/cgi-bin/hgLiftOver). To address these challenges, Chen et al. developed Cyrius, a novel bioinformatics method for all classes of variants and haplotype calling from CYP2D6 in the WGS data (also included in **Table 3**). The tool can overcome CYP2D6 and CYP2D7 homology challenges and work with both GRCh37 and 38 to accurately genotype CYP2D6 with a higher overall concordance rate with true genotypes (99.3%). Compared to Aldy and Stargazer, superior genotyping was demonstrated for both GeT-RM and long-read data, and the application of the method led to improved understanding of CYP2D6 genetic diversity within five ethnic groups. The authors are currently extending the method to genotype other pharmacogenes with a paralog, CYP2A6 and CYP2B6, and plan to apply it to more genes in the future (Chen et al., 2021). Overall, it is useful to be aware of the specifications and the features of each of the tools in order to increase their utility while applying such algorithms to calling different PGx variants out of high throughput sequencing results.

Solutions for the Management of Challenges in Applying the NGS-PGx Tests in the Clinic

Here, we present three main problems which may arise during clinical NGS testing for PGx in everyday practice and discuss solutions.

Firstly, based on the type of panel or other selected approaches, the setup and the initiation of NGS tests (covering PGx markers) in every clinic will require a substantial investment and reimbursement by insurance companies, bioinformatics infrastructure, specific software and computational tools, and professional clinical experts for data interpretation. In addition, validation studies to determine and improve the clinical utility and the validity are essential. Once a positive evaluation has been performed by public and private payers, relevant NGS-derived PGx tests could be considered for implementation in routine clinical practice. Estimated costs of PGx profiling may vary substantially depending on the type of test applied. Is the PGx assessment a pre-emptive NGS test or repurposed findings from diagnostic WES/WGS? Currently, the test coverage and reimbursement are still considered major barriers to routine clinical use. Enhancing physicians' awareness of the type of test to be requested, gaining third-party support, increasing the number of clients through direct-to-consumer genetic testing companies, and decreasing the cost of tests due to advances in diagnostic technologies may play an essential role in bringing the clinical utility of PGx tests to the attention of insurance companies (L Rogers et al., 2020). While many related services are currently limited to reactive single-gene testing, some clinical centers offer routine pre-emptive PGx tests. For example, all patients treated for an active disease at St. Jude Research Hospital are offered PGx testing (www.stjude.org/pg4kds). Recently, Anderson et al. performed a large-scale study in the United States and demonstrated that only a few core pharmacogenes, including CYP2C19, CYP2D6, CYP2C9, VKORC1, UGT1A1, and HLA class I, were covered by the patients' insurance (Anderson et al., 2020).

Secondly, as mentioned previously, the evolutionary conservation is less applicable to the drug-related genes and therefore the conventional computational algorithms have low predictive accuracy when applied to the pharmacogenetic variants. The difficulties with novel and big data interpretation could be overcome by applying combined and optimized calculation tools and algorithms (at least 6-7 of such bioinformatics tools) for allele imputation (see **Appendix 1**) of PGx single- or multi-marker signatures, as well as confirming such genetic variants as predictive for the drug response with more accuracy (Zhou et al., 2019; Tafazoli et al., 2021). However, not all pharmacogenes have this limitation. Indeed, some genes appear relatively free of evolutionary constraints and are highly similar to other genes. This is particularly true for the genes that are involved in the transfer of endogenous substances (i.e., OTC1). Whenever a novel PGx variant is identified in evolutionarily conserved positions, such genes may still benefit

TABLE 4 | Useful databases for PGx results interpretation in the clinical practice.

Database	Main Activities and Features	Link	Reference
PharmGKB	The Pharmacogenomics Knowledgebase is a truly comprehensive and publicly available, online knowledgebase responsible for the aggregation, curation, integration, and dissemination of the knowledge regarding the impact of the human genetic variation on the drug response	https://www.pharmgkb.org/index.jsp	Barbarino et al. (2018)
CPIC	The Clinical Pharmacogenetics Implementation Consortium (CPIC®) is an international consortium to address the clinical implementation of the pharmacogenetic tests by creating, curating, and posting freely available, peer-reviewed, evidence-based, updatable, and detailed gene/drug clinical practice guidelines	https://cpicpgx.org/	Relling and Klein (2011)
DPWG	The Dutch Pharmacogenetics Working Group includes clinical pharmacists, physicians, clinical pharmacologists, clinical chemists, epidemiologists, and toxicologists to develop pharmacogenetics-based therapeutic (dose) recommendations and assist the drug prescribers and the pharmacists by integrating the recommendations into computerized systems for drug prescription and automated medication surveillance	https://www.pharmgkb.org/page/dpwg	JJ Swen et al. (2011)
PharmVar	The Pharmacogene Variation (PharmVar) Consortium is a central repository for the pharmacogene (PGx) variation that focuses on the haplotype structure and the allelic variation. The information in this resource facilitates the interpretation of the pharmacogenetic test results to guide the precision medicine	https://www.pharmvar.org/	Gaedigk et al. (2018)
PMKB	The Precision Medicine Knowledgebase (PMKB) is a project of the Institute of Precision Medicine (IPM) at Weill Cornell Medicine, which is organized to provide information about the clinical cancer variants and the interpretations in a structured way as well as allowing the users to submit and edit the existing entries for the continued growth of the knowledgebase. All changes are reviewed by cancer pathologists	https://pmkb.weill.cornell.edu	Huang et al. (2017)
PharmaADME	An industry-initiated effort launched to develop a consensus, "Core List" of standardized "evidence-based" drug metabolizing (ADME) genetic biomarkers that are broadly applicable to many pharmaceutical clinical trials and FDA drug submissions	http://www.pharmaadme.org/joomla/	pharmaadme and www.pharmaadme.org
Flockhart Table	The website provides a table designed as a hypothesis testing, teaching, and reference tool for the physicians and researchers interested in the drug interactions that are the result of the competition for or effects on the human cytochrome P450 system. The table contains lists of drugs in columns under the designation of specific cytochrome P450 isoforms	https://drug-interactions.medicine.iu.edu/MainTable.aspx	Flockhart and Oesterheld (2000)
SEAPharm	The Southeast Asian Pharmacogenomics Research Network (SEAPharm) established in Asia to enable and strengthen the PGx research among various PGx communities within but not limited to countries in SEA, with the ultimate goal of supporting PGx implementation in the region	–	Chumnumwat et al. (2019)
PGRN	The Pharmacogenomics Research Network, PGRN I–III, was funded from 2000 through 2015 by multiple Institutes and Centers of the NIH. The network catalyzed pharmacogenomics discoveries both nationally and internationally through the conduct of collaborative research focused on the discovery and the translation of the the genetic determinants of the drug response, to enable safer and more effective drug therapies	https://www.pgrn.org/	–
SuperCYP	A comprehensive database on cytochrome P450 enzymes including a tool for analysis of the CYP-drug interactions	https://bioinformatics.charite.de/supercyp/	Preissner et al. (2010)

(Continued on following page)

TABLE 4 | (*Continued*) Useful databases for PGx results interpretation in the clinical practice.

Database	Main Activities and Features	Link	Reference
FDA-Pharmacogenomic	Table of pharmacogenomic biomarkers in drug labeling	https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling	–

from routine predictor tools to indicate their functional impact (Shu et al., 2003). However, in the absence of distinct clinical data, both computational and laboratory models are needed for the genotype-guided drug therapy based on previously unreported genomic variants (Shrestha et al., 2018).

Other PGx specific computational models and algorithms with a high sensitivity and specificity have also been developed for the prediction of the loss of function and/or the functionally neutral variations. The scores obtained with the models could provide quantitative estimation of the impact of different variants on the gene function. A comprehensive analysis of the computational prediction methods and evaluation of the recent progress in the functional interpretation of non-coding variants for drug-metabolizing enzymes and transporters is provided by Zhou and colleagues (Zhou et al., 2018). Once the functionality of a variant is known, the effect on drug pharmacology needs to be estimated. For this, pathway analysis databases as well as DAVID, Human Metabolome Database, String-db, and KEGG could be used to identify the molecular connections between the altered allele(s) in specific genes and the other related genes in the cell. Moreover, newly developed PGx specific tools such as Aldy, Stargazer, Astrolabe, and Cyrius can also help with NGS data processing in the PGx analysis (Klein and Ritchie, 2018; Lee et al., 2019). **Table 4** lists some databases which are useful in interpreting the results of the clinical PGx analysis. We have also recently reviewed the software and the algorithms dedicated to the functional prediction alongside the related mechanism of action in such tools while using the PGx functional analysis (Tafazoli et al., 2021). After finding a potentially strong relationship between the identified variant(s) and the drug response, particular *in-vitro* assessments as well as cell line modifications may be considered for exploring the functional consequences of the altered alleles and diplotypes on the activity of the related protein. However, the latter is not appropriate in clinical use as it increases the turnaround time considerably. As the final step, the clinical association analysis will confirm the connection between the novel variants and the drug response phenotypes in the patients. Needless to say, it is suitable solely for the patient data analysis and not pre-emptive PGx profiling of a healthy individual with no clinically observable phenotype (Ji et al., 2013).

Finally, while well-known and annotated PGx variant(s) can be used immediately in patient care, the clinical translation and utilization of newly introduced variants requires substantial evidence and records of gene-drug interaction as well as phenotyping data. Nevertheless, such data would be stored primarily for the research purposes and the patient may be recontacted for further investigations. Since the prediction of an individual's metabolic status is very important for drug dosage modifications in a clinic, the translation of the sequencing results into phenotype assignment must follow the universal standardized test interpretation approaches. A gene continuum activity score system has been introduced to deal with such situations and may be accepted by reference laboratories and medical centers for converting the genotype data to the clinically actionable recommendations (Hicks et al., 2014). However, to facilitate the incorporation of the high throughput derived PGx reports in the clinical setting, it is necessary to provide the healthcare professionals with more applicable, evidence-based results and employ standardized and updated cohort and case reports (Giri et al., 2019; Krebs and Milani, 2019).

CONCLUSION

The NGS technologies have been used in the PGx research studies for a decade. The rapid development in accessories and supporting bioinformatics tools in addition to the reduced cost and the technological advancement that will allow for testing of a larger number of drug-related genes and biomarkers will result in the widespread use of such methods in various clinical settings. The main challenges are management of identified VUS, a lack of specific variant caller software, poor haplotype phasing, insufficient coverage of some parts of the genome by different platforms, limited capacity to assess variant functionality *in-vitro*, and limited ability to assess functionality through computational approaches. Nevertheless, the application of NGS in PGx testing in the clinical practice is continually increasing, paving the way for new PGx variant discovery and a bright future for pharmacogenomics-guided drug treatment.

AUTHOR CONTRIBUTIONS

AT designed the study, conducted the search for literature, and wrote the entire manuscript. H-JG and JS supervised the study thoroughly and revised and edited the manuscript. WM performed the search for literature and modified the text as well. AK provided the idea and introduced the topic for the manuscript. All authors have read and agreed to the current version of the manuscript.

FUNDING

This article was conducted within the projects which have received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant, agreement no. 754432, and the Polish Ministry of Science and Higher Education and from financial resources for science in 2018–2023 granted for the implementation of an international co-financed project.

REFERENCES

Ammar, R., Paton, T. A., Torti, D., Shlien, A., and Bader, G. D. (2015). Long Read Nanopore Sequencing for Detection of HLA and CYP2D6 Variants and Haplotypes. *F1000Res* 4, 17. doi:10.12688/f1000research.6037.2

Anderson, H. D., Crooks, K. R., Kao, D. P., and Aquilante, C. L. (2020). The Landscape of Pharmacogenetic Testing in a US Managed Care Population. *Genet. Med.* 22, 1247–1253. 10.1038/s41436-020-0788-3.

Arbitrio, M., Di Martino, M., Scionti, F., Agapito, G., Hiram Guzzi, P., and Cannataro, M. (2016). DMETTM (Drug Metabolism Enzymes and Transporters): a Pharmacogenomic Platform for Precision Medicine. *Oncotarget* 5 (33), 54028–54050. doi:10.18632/oncotarget.9927

Ardui, S., Race, V., Zablotskaya, A., Hestand, M. S., Van Esch, H., Devriendt, K., et al. (2017). Detecting AGG Interruptions in Male and Female FMR1 Premutation Carriers by Single-Molecule Sequencing. *Hum. Mutat.* 38, 324–331. doi:10.1002/humu.23150

Barbarino, J. M., Haidar, C. E., Klein, T. E., and Altman, R. B. (2014). PharmGKB Summary. *Pharmacogenetics and genomics* 24, 177–183. doi:10.1097/FPC.0000000000000024

Barbarino, J. M., Whirl-Carrillo, M., Altman, R. B., and Klein, T. E. (2018). PharmGKB: a Worldwide Resource for Pharmacogenomic Information. *Wires Syst. Biol. Med.* 10, e1417. doi:10.1002/wsbm.1417

Bielinski, S. J., Olson, J. E., Pathak, J., Weinshilboum, R. M., Wang, L., Lyke, K. J., et al. (2014). Preemptive Genotyping for Personalized Medicine: Design of the Right Drug, Right Dose, Right Time-Using Genomic Data to Individualize Treatment Protocol. *Mayo Clinic Proc.* 89, 25–33. doi:10.1016/j.mayocp.2013.10.021

Bosma, P. J., Chowdhury, J. R., Bakker, C., Gantla, S., De Boer, A., Oostra, B. A., et al. (1995). The Genetic Basis of the Reduced Expression of Bilirubin UDP-Glucuronosyltransferase 1 in Gilbert's Syndrome. *N. Engl. J. Med.* 333, 1171–1175. doi:10.1056/NEJM199511023331802

Caspar, S. M., Schneider, T., Meienberg, J., and Matyas, G. (2020). Added Value of Clinical Sequencing: WGS-Based Profiling of Pharmacogenes. *Ijms* 21, 2308. doi:10.3390/ijms21072308

Caudle, K. E., Sangkuhl, K., Whirl-Carrillo, M., Swen, J. J., Haidar, C. E., Klein, T. E., et al. (2020). Standardizing CYP 2D6 Genotype to Phenotype Translation: Consensus Recommendations from the Clinical Pharmacogenetics Implementation Consortium and Dutch Pharmacogenetics Working Group. *Clin. Transl Sci.* 13, 116–124. doi:10.1111/cts.12692

Chen, X., Shen, F., Gonzaludo, N., Malhotra, A., Rogert, C., Taft, R. J., et al. (2021). Cyrius: Accurate CYP2D6 Genotyping Using Whole-Genome Sequencing Data. *Pharmacogenomics J.* 21, 251–261. doi:10.1038/s41397-020-00205-5

Choi, J., Tantisira, K. G., and Duan, Q. L. (2019). Whole Genome Sequencing Identifies High-Impact Variants in Well-Known Pharmacogenomic Genes. *Pharmacogenomics J.* 19, 127–135. doi:10.1038/s41397-018-0048-y

Chua, E. W., Cree, S. L., Ton, K. N. T., Lehnert, K., Shepherd, P., Helsby, N., et al. (2016). Cross-Comparison of Exome Analysis, Next-Generation Sequencing of Amplicons, and the iPLEX ADME PGx Panel for Pharmacogenomic Profiling. *Front. Pharmacol.* 7, 1. doi:10.3389/fphar.2016.00001

Chumnumwat, S., Lu, Z. H., Sukasem, C., Winther, M. D., Capule, F. R., Abdul Hamid, A. A. a. t., et al. (2019). Southeast Asian Pharmacogenomics Research Network (SEAPharm): Current Status and Perspectives. *Public health genomics* 22, 132–139. doi:10.1159/000502916

Cohn, I., Paton, T. A., Marshall, C. R., Basran, R., Stavropoulos, D. J., Ray, P. N., et al. (2017). Genome Sequencing as a Platform for Pharmacogenetic Genotyping: a Pediatric Cohort Study. *Npj Genomic Med.* 2, 19. doi:10.1038/s41525-017-0021-8

Cousin, M. A., Matey, E. T., Blackburn, P. R., Boczek, N. J., Mcallister, T. M., Kruisselbrink, T. M., et al. (2017). Pharmacogenomic Findings from Clinical Whole Exome Sequencing of Diagnostic Odyssey Patients. *Mol. Genet. Genomic Med.* 5, 269–279. doi:10.1002/mgg3.283

Cruz-Correa, O. F., León-Cachón, R. B. R., Barrera-Saldaña, H. A., and Soberón, X. (2017). Prediction of Atorvastatin Plasmatic Concentrations in Healthy Volunteers Using Integrated Pharmacogenetics Sequencing. *Pharmacogenomics* 18, 121–131. doi:10.2217/pgs-2016-0072

Dalén, P., Dahl, M.-L., Ruiz, M. L. B., Nordin, J., and Bertilsson, L. (1998). 10-hydroxylation of Nortriptyline in white Persons with 0, 1, 2, 3, and 13 Functional CYP2D6 Genes*. *Clin. Pharmacol. Ther.* 63, 444–452. doi:10.1016/S0009-9236(98)90040-6

Dorado, P., Cáceres, M. C., Pozo-Guisado, E., Wong, M.-L., Licinio, J., and Llerena, A. (2005). Development of a PCR-Based Strategy forCYP2D6genotyping Including Gene Multiplication of Worldwide Potential Use. *Biotechniques* 39, S571–S574. doi:10.2144/000112044

Flockhart, D. A., and Oesterheld, J. R. (2000). Cytochrome P450-Mediated Drug Interactions. *Child. Adolescent Psychiatric Clinics North. America* 9, 43–76. doi:10.1016/s1056-4993(18)30135-4

Gaedigk, A., Ingelman-Sundberg, M., Miller, N. A., Leeder, J. S., Whirl-Carrillo, M., Klein, T. E., et al. (2018). The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clin. Pharmacol. Ther.* 103, 399–401. doi:10.1002/cpt.910

Giri, J., Moyer, A. M., Bielinski, S. J., and Caraballo, P. J. (2019). Concepts Driving Pharmacogenomics Implementation into Everyday Healthcare. *Pgpm* Vol. 12, 305–318. doi:10.2147/PGPM.S193185

Gordon, A. S., Fulton, R. S., Qin, X., Mardis, E. R., Nickerson, D. A., and Scherer, S. (2016). PGRNseq. *Pharmacogenetics and genomics* 26, 161–168. doi:10.1097/FPC.0000000000000202

Guchelaar, H.-J. (2018). Pharmacogenomics, a Novel Section in the European Journal of Human Genetics. *Eur. J. Hum. Genet.* 26, 1399–1400. doi:10.1038/s41431-018-0205-4

Gulilat, M., Lamb, T., Teft, W. A., Wang, J., Dron, J. S., Robinson, J. F., et al. (2019). Targeted Next Generation Sequencing as a Tool for Precision Medicine. *BMC Med. genomics* 12, 81. doi:10.1186/s12920-019-0527-2

Han, S., Park, J., Lee, J., Lee, S., Kim, H., Han, H., et al. (2017). Targeted Next-Generation Sequencing for Comprehensive Genetic Profiling of Pharmacogenes. *Clin. Pharmacol. Ther.* 101, 396–405. doi:10.1002/cpt.532

He, Y., Hoskins, J. M., and Mcleod, H. L. (2011). Copy Number Variants in Pharmacogenetic Genes. *Trends Molecular Medicine* 17, 244–251. doi:10.1016/j.molmed.2011.01.007

Hicks, J., Swen, J., and Gaedigk, A. (2014). Challenges in CYP2D6 Phenotype Assignment from Genotype Data: a Critical Assessment and Call for Standardization. *Cdm* 15, 218–232. doi:10.2174/1389200215666140202215316

Hippman, C., and Nislow, C. (2019). Pharmacogenomic Testing: Clinical Evidence and Implementation Challenges. *Jpm* 9, 40. doi:10.3390/jpm9030040

Huang, L., Fernandes, H., Zia, H., Tavassoli, P., Rennert, H., Pisapia, D., et al. (2017). The Cancer Precision Medicine Knowledge Base for Structured Clinical-Grade Mutations and Interpretations. *J. Am. Med. Inform. Assoc.* 24, 513–519. doi:10.1093/jamia/ocw148

Ikediobi, O., Shin, J., Nussbaum, R., Phillips, K., Translational, U. C. F., Medicine, P. R. O. P., et al. (2009). Addressing the Challenges of the Clinical Application of Pharmacogenetic Testing. *Clin. Pharmacol. Ther.* 86, 28–31. doi:10.1038/clpt.2009.30

Illing, P. T., Purcell, A. W., and Mccluskey, J. (2017). The Role of HLA Genes in Pharmacogenomics: Unravelling HLA Associated Adverse Drug Reactions. *Immunogenetics* 69, 617–630. doi:10.1007/s00251-017-1007-5

Illumina (2020). Infinium® Global Screening Array-24 v3.0 BeadChip. Availableat: https://emea.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/infinium-global-screening-array-data-sheet-370-2016-016.pdf

Ji, Y., Biernacka, J. M., Hebbring, S., Chai, Y., Jenkins, G. D., Batzler, A., et al. (2013). Pharmacogenomics of Selective Serotonin Reuptake Inhibitor Treatment for Major Depressive Disorder: Genome-wide Associations and Functional Genomics. *Pharmacogenomics J.* 13, 456–463. doi:10.1038/tpj.2012.32

Jiang, M., and You, J. H. (2015). Review of Pharmacoeconomic Evaluation of Genotype-Guided Antiplatelet Therapy. *Expert Opin. Pharmacother.* 16, 771–779. doi:10.1517/14656566.2015.1013028

Johnson, J. A., Burkley, B. M., Langaee, T. Y., Clare-Salzler, M. J., Klein, T. E., and Altman, R. B. (2012). Implementing Personalized Medicine: Development of a Cost-Effective Customized Pharmacogenetics Genotyping Array. *Clin. Pharmacol. Ther.* 92, 437–439. doi:10.1038/clpt.2012.125

Ka, S., Lee, S., Hong, J., Cho, Y., Sung, J., Kim, H.-N., et al. (2017). HLAscan: Genotyping of the HLA Region Using Next-Generation Sequencing Data. *BMC bioinformatics* 18, 1–11. doi:10.1186/s12859-017-1671-3

Kalia, S. S., Adelman, K., Adelman, K., Bale, S. J., Chung, W. K., Eng, C., et al. (2017). Recommendations for Reporting of Secondary Findings in Clinical

Exome and Genome Sequencing, 2016 Update (ACMG SF v2.0): a Policy Statement of the American College of Medical Genetics and Genomics. *Genet. Med.* 19, 249–255. doi:10.1038/gim.2016.190

Karnes, J. H., Shaffer, C. M., Bastarache, L., Gaudieri, S., Glazer, A. M., Steiner, H. E., et al. (2017). Comparison of HLA Allelic Imputation Programs. *PLoS One* 12, e0172444. doi:10.1371/journal.pone.0172444

Kawaguchi, S., Higasa, K., Shimizu, M., Yamada, R., and Matsuda, F. (2017). HLA-HD: An Accurate HLA Typing Algorithm for Next-generation Sequencing Data. *Hum. Mutat.* 38, 788–797. doi:10.1002/humu.23230

Klasberg, S., Surendranath, V., Lange, V., and Schöfl, G. (2019). Bioinformatics Strategies, Challenges, and Opportunities for Next Generation Sequencing-Based HLA Genotyping. *Transfus. Med. Hemother* 46, 312–325. doi:10.1159/000502487

Klein, K., Tremmel, R., Winter, S., Fehr, S., Battke, F., Scheurenbrand, T., et al. (2019). A New Panel-Based Next-Generation Sequencing Method for ADME Genes Reveals Novel Associations of Common and Rare Variants with Expression in a Human Liver Cohort. *Front. Genet.* 10, 7. doi:10.3389/fgene.2019.00007

Klein, T. E., and Ritchie, M. D. (2018). PharmCAT: a Pharmacogenomics Clinical Annotation Tool. *Clin. Pharmacol. Ther.* 104, 19–22. doi:10.1002/cpt.928

Krebs, K., and Milani, L. (2019). Translating Pharmacogenomics into Clinical Decisions: Do Not Let the Perfect Be the Enemy of the Good. *Hum. Genomics* 13, 1–13. doi:10.1186/s40246-019-0229-z

L Rogers, S., Keeling, N. J., Giri, J., Gonzaludo, N., Jones, J. S., Glogowski, E., et al. (2020). PARC Report: a Health-Systems Focus on Reimbursement and Patient Access to Pharmacogenomics Testing. *Pharmacogenomics* 21, 785–796. doi:10.2217/pgs-2019-0192

Lam, Y. W. F. (2013). Scientific Challenges and Implementation Barriers to Translation of Pharmacogenomics in Clinical Practice. *ISRN Pharmacol.* 2013, 1–17. doi:10.1155/2013/641089

Larsen, J. B., and Rasmussen, J. B. (2017). Pharmacogenetic Testing Revisited: 5' Nuclease Real-Time Polymerase Chain Reaction Test Panels for Genotyping CYP2D6 and CYP2C19. *Pharmgenomics Pers Med.* 10, 115–128. doi:10.2147/PGPM.S131580

L. Blackburn, H., Schroeder, B., Turner, C., D. Shriver, C., L. Ellsworth, D., and E. Ellsworth, R. (2015). Management of Incidental Findings in the Era of Next-Generation Sequencing. *Cg* 16, 159–174. doi:10.2174/1389202916666150317232930

Lee, E. M. J., Xu, K., Mosbrook, E., Links, A., Guzman, J., Adams, D. R., et al. (2016). Pharmacogenomic Incidental Findings in 308 Families: The NIH Undiagnosed Diseases Program Experience. *Genet. Med.* 18, 1303–1307. doi:10.1038/gim.2016.47

Lee, S.-B., Wheeler, M. M., Patterson, K., Mcgee, S., Dalton, R., Woodahl, E. L., et al. (2019). Stargazer: a Software Tool for Calling star Alleles from Next-Generation Sequencing Data Using CYP2D6 as a Model. *Genet. Med.* 21, 361–372. doi:10.1038/s41436-018-0054-0

Lemieux Perreault, L.-P., Zaïd, N., Cameron, M., Mongrain, I., and Dubé, M.-P. (2018). Pharmacogenetic Content of Commercial Genome-wide Genotyping Arrays. *Pharmacogenomics* 19, 1159–1167. doi:10.2217/pgs-2017-0129

Liau, Y., Maggo, S., Miller, A. L., Pearson, J. F., Kennedy, M. A., and Cree, S. L. (2019). Nanopore Sequencing of the Pharmacogene CYP2D6 Allows Simultaneous Haplotyping and Detection of Duplications. *Pharmacogenomics* 20, 1033–1047. doi:10.1101/57628010.2217/pgs-2019-0080

Madian, A. G., Wheeler, H. E., Jones, R. B., and Dolan, M. E. (2012). Relating Human Genetic Variation to Variation in Drug Responses. *Trends Genetics* 28, 487–495. doi:10.1016/j.tig.2012.06.008

Mantere, T., Kersten, S., and Hoischen, A. (2019). Long-read Sequencing Emerging in Medical Genetics. *Front. Genet.* 10, 426. doi:10.3389/fgene.2019.00426

Marques, S. C., and Ikediobi, O. N. (2010). The Clinical Application of UGT1A1pharmacogenetic Testing: Gene-Environment Interactions. *Hum. Genomics* 4, 1–12. doi:10.1186/1479-7364-4-4-238

Martis, S., Mei, H., Vijzelaar, R., Edelmann, L., Desnick, R. J., and Scott, S. A. (2013). Multi-ethnic Cytochrome-P450 Copy Number Profiling: Novel Pharmacogenetic Alleles and Mechanism of Copy Number Variation Formation. *Pharmacogenomics J.* 13, 558–566. doi:10.1038/tpj.2012.48

Miller, D. T., Lee, K., Chung, W. K., Gordon, A. S., Herman, G. E., Klein, T. E., et al. (2021). ACMG SF v3. 0 List for Reporting of Secondary Findings in Clinical Exome and Genome Sequencing: a Policy Statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.*, 1–10. doi:10.1038/s41436-021-01278-8

Mizzi, C., Peters, B., Mitropoulou, C., Mitropoulos, K., Katsila, T., Agarwal, M. R., et al. (2014). Personalized Pharmacogenomics Profiling Using Whole-Genome Sequencing. *Pharmacogenomics* 15, 1223–1234. doi:10.2217/pgs.14.102

Nakkam, N., Konyoung, P., Kanjanawart, S., Saksit, N., Kongpan, T., Khaeso, K., et al. (2018). HLA Pharmacogenetic Markers of Drug Hypersensitivity in a Thai Population. *Front. Genet.* 9, 277. doi:10.3389/fgene.2018.00277

Numanagić, I., Malikić, S., Pratt, V. M., Skaar, T. C., Flockhart, D. A., and Sahinalp, S. C. (2015). Cypiripi: Exact Genotyping of CYP2D6 Using High-Throughput Sequencing Data. *Bioinformatics* 31, i27–34. doi:10.1093/bioinformatics/btv232

Numanagić, I., Malikić, S., Ford, M., Qin, X., Toji, L., Radovich, M., et al. (2018). Allelic Decomposition and Exact Genotyping of Highly Polymorphic and Structurally Variant Genes. *Nat. Commun.* 9, 1–11. doi:10.1038/s41467-018-03273-1

Oscarson, M., Hidestrand, M., Johansson, I., and Ingelman-Sundberg, M. (1997). A Combination of Mutations in the CYP2D6*17(CYP2D6Z) Allele Causes Alterations in Enzyme Function. *Mol. Pharmacol.* 52, 1034–1040. doi:10.1124/mol.52.6.1034

Owen, R. P., Gong, L., Sagreiya, H., Klein, T. E., and Altman, R. B. (2010). VKORC1 Pharmacogenomics Summary. *Pharmacogenetics and genomics* 20, 642–644. doi:10.1097/FPC.0b013e32833433b6

Pharmaadmeand Www.Pharmaadme.Org .

Preissner, S., Kroll, K., Dunkel, M., Senger, C., Goldsobel, G., Kuzman, D., et al. (2010). SuperCYP: a Comprehensive Database on Cytochrome P450 Enzymes Including a Tool for Analysis of CYP-Drug Interactions. *Nucleic Acids Res.* 38, D237–D243. doi:10.1093/nar/gkp970

Price, M. J., Carson, A. R., Murray, S. S., Phillips, T., Janel, L., Tisch, R., et al. (2012). First Pharmacogenomic Analysis Using Whole Exome Sequencing to Identify Novel Genetic Determinants of Clopidogrel Response Variability: Results of the Genotype Information and Functional Testing (GIFT) Exome Study. *J. Am. Coll. Cardiol.* 59, E9. doi:10.1016/S0735-1097(12)60010-2

Rasmussen-Torvik, L. J., Almoguera, B., Doheny, K. F., Freimuth, R. R., Gordon, A. S., Hakonarson, H., et al. (2017). Concordance between Research Sequencing and Clinical Pharmacogenetic Genotyping in the eMERGE-PGx Study. *J. Mol. Diagn.* 19, 561–566. doi:10.1016/j.jmoldx.2017.04.002

Reisberg, S., Krebs, K., Lepamets, M., Kals, M., Mägi, R., Metsalu, K., et al. (2019). Translating Genotype Data of 44,000 Biobank Participants into Clinical Pharmacogenetic Recommendations: Challenges and Solutions. *Genet. Med.* 21, 1345–1354. doi:10.1038/s41436-018-0337-5

Relling, M. V., and Klein, T. E. (2011). CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin. Pharmacol. Ther.* 89, 464–467. doi:10.1038/clpt.2010.279

Sabater, A., Ciudad, C., Cendros, M., Dobrokhotov, D., and Sabater-Tobella, J. (2019). G-Nomic: a New Pharmacogenetics Interpretation Software. *Pgpm* Vol. 12, 75–85. doi:10.2147/PGPM.S203585

Saminathan, R., Bai, J., Sadrolodabaee, L., Karthik, G. M., Singh, O., Subramaniyan, K., et al. (2010). VKORC1 Pharmacogenetics and Pharmacoproteomics in Patients on Warfarin Anticoagulant Therapy: Transthyretin Precursor as a Potential Biomarker. *PLoS One* 5, e15064. doi:10.1371/journal.pone.0015064

Sangkuhl, K., Whirl-Carrillo, M., Whaley, R. M., Woon, M., Lavertu, A., Altman, R. B., et al. (2020). Pharmacogenomics Clinical Annotation Tool (Pharm CAT). *Clin. Pharmacol. Ther.* 107, 203–210. doi:10.1002/cpt.1568

Santos, M., Niemi, M., Hiratsuka, M., Kumondai, M., Ingelman-Sundberg, M., Lauschke, V. M., et al. (2018). Novel Copy-Number Variations in Pharmacogenes Contribute to Interindividual Differences in Drug Pharmacokinetics. *Genet. Med.* 20, 622–629. doi:10.1038/gim.2017.156

Schärfe, C. P. I., Tremmel, R., Schwab, M., Kohlbacher, O., and Marks, D. S. (2017). Genetic Variation in Human Drug-Related Genes. *Genome Med.* 9, 117. doi:10.1186/s13073-017-0502-5

Schwarz, U. I., Gulilat, M., and Kim, R. B. (2019). The Role of Next-Generation Sequencing in Pharmacogenetics and Pharmacogenomics. *Cold Spring Harb Perspect. Med.* 9, a033027. doi:10.1101/cshperspect.a033027

Shrestha, S., Zhang, C., Jerde, C. R., Nie, Q., Li, H., Offer, S. M., et al. (2018). Gene-Specific Variant Classifier (DPYD-Varifier) to Identify Deleterious Alleles of Dihydropyrimidine Dehydrogenase. *Clin. Pharmacol. Ther.* 104, 709–718. doi:10.1002/cpt.1020

Shu, Y., Leabman, M. K., Feng, B., Mangravite, L. M., Huang, C. C., Stryke, D., et al. (2003). Evolutionary Conservation Predicts Function of Variants of the Human Organic Cation Transporter, OCT1. *Proc. Natl. Acad. Sci.* 100, 5902–5907. doi:10.1073/pnas.0730858100

Sivadas, A., Salleh, M. Z., Teh, L. K., and Scaria, V. (2017). Genetic Epidemiology of Pharmacogenetic Variants in South East Asian Malays Using Whole-Genome Sequences. *Pharmacogenomics J.* 17, 461–470. doi:10.1038/tpj.2016.39

Suarez-Kurtz, G., and Parra, E. J. (2018). "Population Diversity in Pharmacogenetics: a Latin American Perspective," in *Advances in Pharmacology* (Elsevier), 133–154. doi:10.1016/bs.apha.2018.02.001

Svidnicki, M. C. C. M., Zanetta, G. K., Congrains-Castillo, A., Costa, F. F., and Saad, S. T. O. (2020). Targeted Next-Generation Sequencing Identified Novel Mutations Associated with Hereditary Anemias in Brazil. *Ann. Hematol.* 99, 955–962. doi:10.1007/s00277-020-03986-8

Swen, J. J., Nijenhuis, M., de Boer, A., Grandia, L., Maitland-van der Zee, A. H., Mulder, H., et al. (2011). Pharmacogenetics: From Bench to Byte- an Update of Guidelines. *Clin. Pharmacol. Ther.* 89, 662–673. doi:10.1038/clpt.2011.34

Tafazoli, A., Wawrusiewicz-Kurylonek, N., Posmyk, R., and Miltyk, W. (2021). Pharmacogenomics, How to Deal with Different Types of Variants in Next Generation Sequencing Data in the Personalized Medicine Area. *Jcm* 10, 34. doi:10.3390/jcm10010034

Taylor, C., Crosby, I., Yip, V., Maguire, P., Pirmohamed, M., and Turner, R. M. (2020). A Review of the Important Role of CYP2D6 in Pharmacogenomics. *Genes* 11, 1295. doi:10.3390/genes11111295

Thermofisher.Com/Pharmacoscan (2018). Verification of Buccal Swab and Saliva Sample Types for PharmacoScan Solution.

Twesigomwe, D., Wright, G. E., Drögemöller, B. I., Da Rocha, J., Lombard, Z., and Hazelhurst, S. (2020). A Systematic Comparison of Pharmacogene star Allele Calling Bioinformatics Algorithms: a Focus on CYP2D6 Genotyping. *NPJ Genomic Med.* 5, 1–11. doi:10.1038/s41525-020-0135-2

Twist, G. P., Gaedigk, A., Miller, N. A., Farrow, E. G., Willig, L. K., Dinwiddie, D. L., et al. (2016). Constellation: a Tool for Rapid, Automated Phenotype Assignment of a Highly Polymorphic Pharmacogene, CYP2D6, from Whole-Genome Sequences. *Npj Genomic Med.* 1, 1–10. doi:10.1038/npjgenmed.2015.7

Twist, G. P., Gaedigk, A., Miller, N. A., Farrow, E. G., Willig, L. K., Dinwiddie, D. L., et al. (2017). Erratum: Constellation: a Tool for Rapid, Automated Phenotype Assignment of a Highly Polymorphic Pharmacogene, CYP2D6, from Whole-Genome Sequences. *Npj Genomic Med.* 2, 16039. doi:10.1038/npjgenmed.2016.39

van der Lee, M., Allard, W. G., Bollen, S., Santen, G. W. E., Ruivenkamp, C. A. L., Hoffer, M. J. V., et al. (2020a). Repurposing of Diagnostic Whole Exome Sequencing Data of 1,583 Individuals for Clinical Pharmacogenetics. *Clin. Pharmacol. Ther.* 107, 617–627. doi:10.1002/cpt.1665

Van Der Lee, M., Kriek, M., Guchelaar, H.-J., and Swen, J. J. (2020b). Technologies for Pharmacogenomics: A Review. *Genes* 11, 1456. doi:10.3390/genes11121456

Walczak, M., Skrzypczak-Zielinska, M., Plucinska, M., Zakerska-Banaszak, O., Marszalek, D., Lykowska-Szuber, L., et al. (2019). Long-range PCR Libraries and Next-Generation Sequencing for Pharmacogenetic Studies of Patients Treated with Anti-TNF Drugs. *Pharmacogenomics J.* 19, 358–367. doi:10.1038/s41397-018-0058-9

Wennerholm, A., Dandara, C., Sayi, J., Svensson, J. O., Abdi, Y. A., Ingelman-Sundberg, M., et al. (2002). The African-specific CYP2D6*17 Allele Encodes an Enzyme with Changed Substrate Specificity. *Clin. Pharmacol. Ther.* 71, 77–88. doi:10.1067/mcp.2002.120239

Westbrook, M. J., Wright, M. F., Van Driest, S. L., Mcgregor, T. L., Denny, J. C., Zuvich, R. L., et al. (2013). Mapping the Incidentalome: Estimating Incidental Findings Generated through Clinical Pharmacogenomics Testing. *Genet. Med.* 15, 325–331. doi:10.1038/gim.2012.147

Wouden, C. H., Van Rhenen, M. H., Jama, W. O. M., Ingelman-Sundberg, M., Lauschke, V. M., Konta, L., et al. (2019). Development of the PG x-Passport: A Panel of Actionable Germline Genetic Variants for Pre-Emptive Pharmacogenetic Testing. *Clin. Pharmacol. Ther.* 106, 866–873. doi:10.1002/cpt.1489

Xie, C., Yeo, Z. X., Wong, M., Piper, J., Long, T., Kirkness, E. F., et al. (2017). Fast and Accurate HLA Typing from Short-Read Next-Generation Sequence Data with xHLA. *Proc. Natl. Acad. Sci. USA* 114, 8059–8064. doi:10.1073/pnas.1707945114

Yang, Y., Botton, M. R., Scott, E. R., and Scott, S. A. (2017). Sequencing theCYP2D6gene: from Variant Allele Discovery to Clinical Pharmacogenetic Testing. *Pharmacogenomics* 18, 673–685. doi:10.2217/pgs-2017-0033

Zhou, Y., Fujikura, K., Mkrtchian, S., and Lauschke, V. M. (2018). Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data. *Front. Pharmacol.* 9, 1437. doi:10.3389/fphar.2018.01437

Zhou, Y., Mkrtchian, S., Kumondai, M., Hiratsuka, M., and Lauschke, V. M. (2019). An Optimized Prediction Framework to Assess the Functional Impact of Pharmacogenetic Variants. *Pharmacogenomics J.* 19, 115–126. doi:10.1038/s41397-018-0044-2

Zidan, A. M., Saad, E. A., Ibrahim, N. E., Mahmoud, A., Hashem, M. H., and Hemeida, A. A. (2020). PHARMIP: An Insilico Method to Predict Genetics that Underpin Adverse Drug Reactions. *MethodsX* 7, 100775. doi:10.1016/j.mex.2019.100775

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

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

APPENDIX 1: MINI-GLOSSARY OF THE NGS TERMINOLOGIES USED IN THIS ARTICLE

Targeted sequencing	Sequencing of specific parts of the genome or sets of genes (multiple genes) at once
Whole-exome sequencing	Sequencing of all exonic (protein-coding) regions of the genome. It also includes some important flanking sequences
Whole-genome sequencing	Sequencing of the entire genome of an organism, including all the non-coding and coding parts in addition to the mitochondrial DNAs. It is the most comprehensive sequencing method among others
Long-read sequencing	Newer sequencing methods (also called third-generation technologies) with the ability to read and produce long sequences of DNA between 10,000 and 100,000 base pairs at one runtime. The method is useful, particularly for the structural variant detection and haplotype phasing
Gene panel	A specific set of genes selected for particular analysis purposes as well as sequencing methods or disease-specific gene profiling
PGx panel	A specific set of pharmacogenes or drug target genes selected for particular analysis purposes
Coverage	The number of times a portion of the genome is sequenced in a sequencing reaction. Frequently expressed as “depth of coverage” and numerically as 1X, 2X, 3X, etc.
Depth of coverage	See above
VCF file	Variant calling format is a standard variant reporting format which was invented during the 1,000 Genomes project. Such files display the genomic variants with their coordinates in the NGS results
Secondary findings	Unrelated genomic variants to the primary purpose of the test revealed during a sequencing run
Haplotype phasing	Determination of paternal or maternal origins (inheritance) of each chromosome while putting into haplotypes. In this way, the researchers can assign the alleles to the paternal and maternal chromosomes and obtain a comprehensive picture of genomic variants for the specific haplotype Statistical estimation of the haplotypes from the genotyping data is also called haplotype phasing
Allele imputation	A statistical method for inferring the genotypes that are not directly measured. Estimation of unobserved genotype, including genetic markers from known haplotype or reference genotype. Particularly beneficial in GWAS studies
VUS	A genetic variant for which the association with a specific phenotype cannot be determined definitively

Review article

Pharmacogenomics, How to Deal with Different Types of Variants in Next Generation Sequencing Data in the Personalized Medicine Area

Alireza Tafazoli, Natalia Wawrusiewicz-Kurylonek, Renata Posmyk, and Wojciech Miltyk*

Journal of Clinical Medicine

DOI: 10.3390/jcm10010034

IF: 4.964
MEiN: 140

Journal of
Clinical Medicine

MDPI

Review

Pharmacogenomics, How to Deal with Different Types of Variants in Next Generation Sequencing Data in the Personalized Medicine Area

Alireza Tafazoli [1,2], Natalia Wawrusiewicz-Kurylonek [3], Renata Posmyk [3] and Wojciech Miltyk [1,*]

[1] Department of Analysis and Bioanalysis of Medicines, Faculty of Pharmacy with the Division of Laboratory Medicine, Medical University of Białystok, 15-089 Białystok, Poland; Alireza.Tafazoli@umb.edu.pl
[2] Clinical Research Centre, Medical University of Białystok, 15-276 Bialystok, Poland
[3] Department of Clinical Genetics, Medical University of Białystok, 15-089 Białystok, Poland; natalia.kurylonek@gmail.com (N.W.-K.); rposmyk@gmail.com (R.P.)
* Correspondence: wojciech.miltyk@umb.edu.pl; Tel.: +48-857485845

Abstract: Pharmacogenomics (PGx) is the knowledge of diverse drug responses and effects in people, based on their genomic profiles. Such information is considered as one of the main directions to reach personalized medicine in future clinical practices. Since the start of applying next generation sequencing (NGS) methods in drug related clinical investigations, many common medicines found their genetic data for the related metabolizing/shipping proteins in the human body. Yet, the employing of technology is accompanied by big obtained data, which most of them have no clear guidelines for consideration in routine treatment decisions for patients. This review article talks about different types of NGS derived PGx variants in clinical studies and try to display the current and newly developed approaches to deal with pharmacogenetic data with/without clear guidelines for considering in clinical settings.

Keywords: pharmacogenomics; NGS variants; personalized medicine

Citation: Tafazoli, A.; Wawrusiewicz-Kurylonek, N.; Posmyk, R.; Miltyk, W. Pharmacogenomics, How to Deal with Different Types of Variants in Next Generation Sequencing Data in the Personalized Medicine Area. *J. Clin. Med.* **2021**, *10*, 34. https://dx.doi.org/10.3390/jcm10010034

Received: 30 November 2020
Accepted: 22 December 2020
Published: 24 December 2020

1. Introduction: Pharmacogenomics and High Throughput Sequencing Methods

It has been reported for decades that different drugs show different responses and efficacy in diverse individuals or populations. Investigations proved that part of this diversity (20–30%) is because of genetic background and, more precisely, the inheritance of various alleles and variants in genes for drug-metabolizing and transporting (pharmacogenes) or drug target molecules [1]. Pharmacogenetics is the term for the knowledge of diverse drug responses and effects in people, based on their single genes on the genomic profiles. When a group of genes (multiple genes), or whole genome, and other influential genomic events, such as epigenetics will be addressed at once for such investigations, the phrase would be replaced by pharmacogenomics (PGx). Since the starting of employing high throughput sequencing methods, especially next generation sequencing (NGS) technologies, in addition to some comprehensive orthogonal tests, such as genome-wide single nucleotide polymorphism (SNP) arrays in clinical investigations and practice, numerous genetic variants have been introduced in drug-related genes in the human body. Today, close to 100 variants in each people in more than 900 of such genes are mentioned in literature, and the number is increasing continuously [2,3]. There is no doubt that the NGS methods played a significant role in the identification of PGx variants in a clinical research setting and used in the prediction of the response to or adverse effects of drugs, which result in the calculation or estimation of appropriate drug dosage for patients. According to the patient's responses, the drug outcome could be defined as efficient, inefficient, toxic, and resistant. All of these categories mostly arise from the interaction between the products of many genes in a cellular pathway or between the genes and environmental factors. Hence, genotype-specific therapy could bring huge benefits for drug safety and efficacy in patients

in addition to time and cost reduction of treatment approaches for them [4]. The trends led to the practice of personalized therapy and precision medicine implementation in clinical centers. The explosion of examples in the field of pre-emptive and/or patient genotyping shows the true advantages of high throughput sequencing technologies in the PGx area [5–8]. However, despite the common belief between the physicians and general practitioners in the effects of the genetic landscape on diverse drug responses, if they asked that they order the PGx tests for their patients, less than 15% will answer positively. This is mostly because of the lack of clear guidelines and sufficient clinical evidence for many functional genetic variants (FGVs) in drug-related genes (*FGVs or actionable genetic variants are those alterations in genome, with at least one report for introducing the effects on drug safety and/or efficacy in people. Moreover, the variants found in the research area with strong potential effects on drugs could be considered as FGVs during prescription. However, the latter needs clinical evidence to be influential on treatment decisions by physicians*). Furthermore, the poor knowledge and background of PGx and the different related alleles and variants for many healthcare professionals may directly affect their desire to order the tests.

Yet, several rare and uncommon FGVs can be detected through the PGx tests in both clinical and research areas, especially when comprehensive and high capacity methods, such as NGS, have been utilized [9]. Moreover, it is necessary to distinguish the definition of FGVs and/or uncharacterized variants, such as variants with unknown clinical significance in two distinct genomic medicine areas, PGx, and medical genetics. Although the two concepts are usually mixed and many PGx variants are covered in the medical genetics zone, the first one mostly emphasizes those variants with an impact on pharmacological treatments, while the second group of variants is considered the genetic variations with pathogenicity effects in the human body. For a PGx variant, it might show an interaction with drug dosage modifications or not, but the functional and clinical consequences of a genetic variant may be unknown (does it have pathological consequences?) or well known (it has or not pathological consequences). However, both types of variants will be addressed as the same in NGS primary data analysis steps. To deal with the different genetic variants in PGx profiling of individuals, this review article reviews various NGS derived biomarkers and the possible approaches to use or consider them during the medicine prescription. Those PGx variants with no clear guidelines will be focused on more.

2. Different Types of Variants and Their Classifications in Clinical Pharmacogenomics

Both common and rare alleles are demonstrated as the functional biomarkers in PGx clinical practice. Low frequency and rare variants have been shown by 1–5% and lower than 1% minor allele frequency (MAF), respectively, in populations. Moreover, they proved to be very population-specific and the causative elements for diverse drug responses in alternative ethnic groups [10,11]. NGS methods revolutionized the detection of any type of variants in different aspects of genome analysis and profiling, as well as pharmacogenetics and genomic studies. Such investigations reported that most of the FGVs in the clinical PGx setting are Single Nucleotide Variations (SNVs). However, structural variants (SVs), such as Copy Number Variation (CNVs), small Insertion–Deletions (InDels), tandem-substitutions, and the deletion of entire exons are also identified as effective variants in drug responses [12,13]. In addition to wild-type alleles, the functional outcome for each of these variants may cause the individuals to fall into four main groups of responders including poor, intermediate, extensive, and ultra-rapid metabolizers.

Currently, core web-based resources for clinical PGx annotations include Pharmacogenomics Knowledge Base (PharmGKB), the Clinical Pharmacogenetics Implementation Consortium (CPIC), the Pharmacogenomics Research Network (PGRN), and Dutch Pharmacogenetics Working Group (DPWG). These are considered as reference databases that provide information about how human genetic variations affect response to medications. All of the confirmed data about clinically actionable gene-drug associations and genotype-phenotype relationships are sorted properly and available as a guide for personalized medicine implementation by healthcare professionals. However, other modules, such as

PharmVar, FINDbase, SuperCYP, SEAPharm, etc. could also be applied when a specific type of gene or drug was on the desk. Nevertheless, according to PGx reference organizations (PharmGKB, CPIC-PGRN, and DPWG), all the diagnosed alleles and variants in a gene-drug interaction, based on the number of published studies and clinical evidence, will be classified in various types of level with clear explanations for each of them (Table 1). However, CPIC has also introduced a new categorization system for PGx level in more detail (Table 2). Generally, different levels of clinical relevance for PGx variants and/or gene-drug pairs will be assigned by the reference entities. All of them have their processes to assign the levels and prioritize approaches for providing the related guidelines. Meanwhile, some recommendations are related to each other (CPIC and PharmGKB) and the others go through it independently (DPWG). For example, the clinical pharmacogenetics implementation consortium (CPIC) allocates the levels for a variant in a gene-drug pair, based on three major criteria from PharmGKB clinical annotation levels of evidence and PGx level for Food and Drug Administration (FDA)-approved drug labels and also if it is nominated to CPIC for consideration. Only those gene/drug pairs that have been the subject of guidelines have had sufficient in-depth review of evidence to provide definitive CPIC level assignments. CPIC also use other considerations for assignment of CPIC level through some essential questions, containing the information of prescribing actionability, the severity of the clinical consequences for ignoring the genetic tests, already subjected gene to other CPIC guidelines, availability of genetic test for the gene, high-risk genetic variants, etc. [14,15]. PharmGKB also creates genotype-based summaries describing the phenotypic impact of the variant and provides the PGx levels from 1A to 4 in combination with four instructive labels as “Testing required”, “Testing recommended”, “Actionable PGx”, and “Informative PGx” via literature reviews while considering population size and statistical significance. The labels state different considerations for the drugs, based on gene/protein/chromosomal variants or phenotypes, and conclude the necessity of pre-emptive genetic testing for genotype/phenotype correlation assays and showing the potential changes in efficacy, dosage, metabolism, or toxicity [16,17]. Finally, the Dutch Pharmacogenetics working group (DPWG) uses the drug-gene interaction outcomes to providing the clinical relevance levels, where the AA is the lowest impact and F is the highest one. The impacts are categorized, based on adverse drug events, decreased therapeutic response, and other clinical effects, result in the allocation of specific scores from 1–7 derived from national cancer institute (NCI) common toxicity criteria and 0–4 level of evidence of gene–drug interaction in the literature [18].

Table 1. Different levels of clinical relevance for pharmacogenomics (PGx) variants in reference organizations.

Reference Organization	*PGx Level*	*Summary of Description*	*Reference*
PharmGKB	***1A***	Variants in this level are annotated and have a clear and endorsed guideline while showing a strong role in gene-drug interactions.	[19]
	1B	Annotated variant with strong evidence in the literature. Gene-drug association shows strong effects.	
	2A	The annotated variant is in a VIP *, so functional significance is more likely.	
	2B	Annotated variant but in moderate evidence of an association. There is no reliable replicated study in form of statistical significance or well-designed in size.	
	3	Annotated variant in a single study or multiple studies with no similar associations between the variant and the drug.	
	4	Annotated variant but in a case report and non-significant study or just in an in-vitro assay.	
CPIC	***A***	Variants in this level oblige a change in related drug prescription. Strong clinical evidence and genotype-phenotype correlations exist.	[20]
	B	Evidence is weak for the variant but still genotyping may be useful for alternative prescribing.	
	C	Different levels of evidence are mentioned in various publications for the variant. No prescribing actions are recommended. Mostly suitable for genes that are commonly included in clinical or DTC ** tests.	
	D	Weak evidence and conflicting data are introduced for the variant. Clinical actionability is unclear. No prescribing actions are recommended.	
DPWG	***AA***	Variants with no significant clinical or kinetic effects.	[21]
	A	Variants with minor clinical effects and kinetic effects.	
	B	Variants with mild clinical effects.	
	C	Variants with moderate clinical effects.	
	D	Variants with stronger clinical effects than level C.	
	E	Variants with severe clinical effects as the failure of lifesaving therapy or life-threatening complications.	
	F	Variants with most severe clinical effects, death is anticipated.	
	***		
	4	There are good quality published studies for the variant/gene.	
	3	There are moderate quality published studies for the variant/gene.	
	2	Well documented case reports exist for the variant/gene.	
	1	Published incomplete case reports for the variant/gene.	
	0	Data on file.	
	???	No evidence.	

* VIP: very important pharmacogene, ** DTC: direct to consumer, *** Separate the two different levels definitions of the DPWG.

Table 2. Clinical Pharmacogenetics Implementation Consortium (CPIC) new level of clinical relevance for gene/drug interactions.

Cpic Level	*Clinical Context*	*Level of Evidence*	*Strength of Recommendation*
A	Genetic information should be used to change the prescribing of the affected drug.	The preponderance of the evidence is high or moderate in favor of changing prescribing.	At least one moderate or strong action (change in prescribing) is recommended.
A/B	Preliminary review indicates it is likely that the definitive CPIC level will be either A or B.	Full evidence review is needed to assess the level of evidence, but prescribing actionability is likely.	Full review by expert guideline group to assign strength of recommendation.
B	Genetic information could be used to change prescribing of the affected drug because alternative therapies/dosing are extremely likely to be as effective and as safe as non-genetically based dosing.	The preponderance of the evidence is weak with little conflicting data.	At least one optional action (change in prescribing) is recommended.
B/C	Preliminary review indicates it is likely that the definitive CPIC level will be either B or C.	Prescribing actionability based on genetics is not clear without further evidence review.	Full review by expert guideline group to assess the strength of recommendation.
C	There are published studies at varying levels of evidence, some with mechanistic rationale, but no prescribing actions are recommended because (a) dosing based on genetics makes no convincing difference; (b) alternatives are unclear, possibly less effective, more toxic, or otherwise impractical; or (c) few published studies or mostly weak evidence and clinical actions are unclear. Most important for genes that are subject to other CPIC guidelines or genes that are commonly included in clinical or DTC tests.	Evidence levels can vary.	No prescribing actions are recommended.
C/D	Preliminary review indicates it is likely that the definitive CPIC level will be either C or D.	Evidence levels can vary.	No prescribing actions are recommended.
D	There are few published studies, clinical actions are unclear, little mechanistic basis, mostly weak evidence, or substantial conflicting data. If the genes are not widely tested clinically, evaluations are not needed. Criteria for "widely tested" includes: 1) College of American Pathologists (CAP) proficiency testing is available; 2) gene is in disease-specific panels (e.g., pain, psychiatric, cancer, etc.); or 3) evidence exists for implementation of the gene into clinical practice (CPIC member feedback, publications, etc.).	Evidence levels can vary.	No prescribing actions are recommended.

Adopted from cpicpgx.org/.

Regarding the abovementioned level of classification for the identified variants, the utilization of NGS platforms for clinical PGx tests brings various types of alleles, which after confirmation and validation processes could be categorized as functional/potential effective variants, fall into "five groups of (1) annotated variants with the clear guideline (i.e., rs1057910 in *CYP2C9* and rs9923231 in *VKORC1* genes for Warfarin). (2) Annotated variants with no clinical guideline (i.e., rs6166 in *FSHR* gene for urofollitropin). (3) Variants with annotation or guidelines for other drugs (i.e., rs9322335 in *ESR1* gene for letrozole while the gene is studying and considered as the estrogen receptor and target molecule for

Clomifene). (4) Non-pharmacogenetically annotated variants (i.e., different clinical related variants in *AR* gene as an important target molecule for infertility drugs). And (5) Variants of unknown significance (VUS). The next part will focus on different approaches for such variant interpretation and curation in clinical practice.

3. Approaches to Dealing with Diverse Pharmacogenomics Variants

To finding any clinical relevance for different groups of PGx variants from the sequencing platforms, standard algorithms, and procedures are introduced by the reference sources (Figure 1). These are the recommendations that indicate the approaches for decoding or predicting the variant functions and the related phenotypes as the diverse drug responses in individuals [22]. From the previous section, group 1 is considered as straightforward, actionable variants in gene-drug pairs with direct prescription recommendations for applying in routine clinical practice. Group 2 are the alleles, consisting of the most common types of identified variants during diagnostic procedures for PGx tests. As the PharmGKB included 19,028 variant annotations, most of the identified markers will fall into this group. Here, the number of clinical evidence in addition to statistical signification (i.e., number of patients in cohort studies) and types of the publications, if they are strong genome-wide association study, well designed replicated report, case report, non-significant study, or only an in-vitro study, would be the important factors for clinical consideration and decisions [23]. The other common scenario for the sequencing results of a pharmacogenetic screening test could be found in group 3, which are variants with the recommendations but not for the researchers/clinicians targeted drugs. Generally, if the related gene is introduced as a very important pharmacogene (VIP) in PGx databases, it is mostly well documented so the related cellular pathways must be analyzed thoroughly. Then the caution and consideration before dosage adjustment are suggested for more accurate implementation of personalized medicine in the clinic [24]. If there is a lack of such documents, more confirmation and validation assessments are necessary before any concerns for the patient's prescription. Replicate tests in target drugs in such situations consist of various approaches, from looking for the same result in same/different ethnic groups to implementation of laboratory confirmation tests. However, alternative approaches have also been introduced for PGx findings validation, if replication studies for gene-drug interactions proved to be difficult and costly for some cases [25]. In the end, consulting with gene experts or experienced clinical pharmacologist in the gene–drug interaction field is necessary. So far, reference databases have explained the approaches to deal with variants in group one to three. However, many genetic variations may be classified in group 4, which is introduced as disease-associated biomarkers and placed into the different genomic databases, such as ClinVar, dbGaP, HapMap, gnomAD, COSMIC, etc. (as causative or pathogenic variants), but there is no PGx report for them. This is mostly happening during more comprehensive genomic profiling of individuals for decoding any PGx markers. In such a situation, the first step could be the evaluation of the gene, if it is introduced as drug related in literature and databases before. The positive result may follow the approaches for group 3 as well. If there is any, also clinical assays would help provide evidence in both groups 2 and 3 of variants during the clinical decision making.

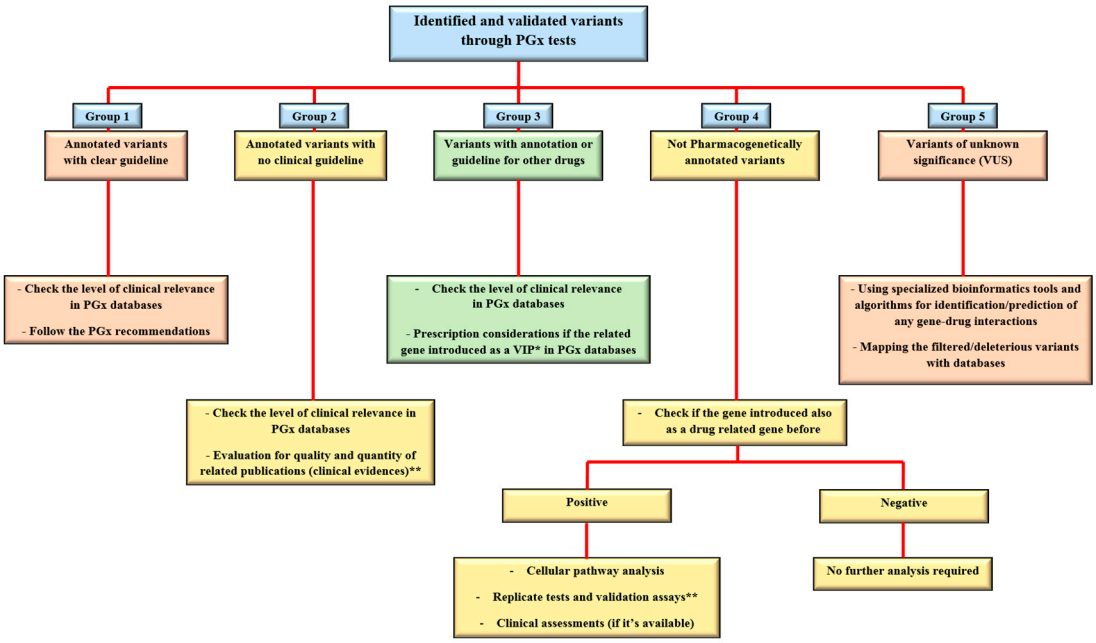

Figure 1. Approaches to deal with different types of PGx variants in clinical centers. After the identification and doing the confirmation tests on a PGx related variant, it could be categorized in one of the main five groups of annotated with PGx guideline, annotated without a guideline, informative for other drugs, not PGx annotated, or variants of unknown clinical significance (VUS). For the annotated variants, checking the level of clinical relevance (Table 1 of the current paper) is the first task to do. Bioinformatics tools are also supporting the analysis of not only VUS but also other types of variants in each group. *Examples for groups 1–5 with explanations are provided in the main text. * VIP: very important pharmacogene. ** see the text for more details.*

The last types of variants (group 5) are the novel and unreported variations in databases (ClinVar, HGMD, PharmGKB), but found in a PGx test mostly through comprehensive methods, such as whole exome or whole genome sequencing (WES and WGS), with no clue for their function in causing a particular phenotype. Moreover, incidental findings (IFs) are the group of known variations, but not related to specifically investigated phenotype, and accidentally revealed during a sequencing test. Both the VUS (novel variants) and IFs will be manageable with higher accuracy by the combined usage of highly specialized bioinformatics pipelines to find any possible interaction with drug responses in patients. IFs are mostly displayed as the annotated functional drug-related variants in pharmacogenes and potentially useful markers if the appropriate genomic analysis and accurate genotype–phenotype correlations are performed subsequently [26]. We will address this topic in detail in the following section.

4. Approaches to Dealing with Novel Pharmacogenomics Variants

As the majority of revealed variants through implementation of broad range high throughput sequencing tests could be categorized in group 4 and 5 (the most challenging groups), the process of identifying clinically relevant PGx variants from complex genomic data mostly concerns about the detection of any potential FGVs in these two categories. The procedures usually start with digging the variant call format (VCF) file for filtration of variants and selection of those alterations, which come from drug-related genes. Based on the employed sequencer machine and the selected platform for PGx data clinical assessment, different types of variants are available in subsequent result analysis (SNVs and/or CNVs from coding and noncoding/regulatory parts of the genome). Routine silico analysis is considered for filtration of NGS derived pharmacovariants data at the first step (including the quality assessments, segregation studies, zygosity mapping, and allele phasing, etc.). Next, the selected variants go for pathogenicity and functional annotation analysis

through the utilization of prediction algorithms in both common (i.e., *SIFT*, *PolyPhen2*, *MutationTaster*) and PGx dedicated tools (i.e., *Stargazer*, *Aldy*, *Astrolabe*) [27–29]. As the final stage, computational and in-vitro confirmation studies can aid in the identification of prediction's sensitivity, specificity, and accuracy level. This is usually implemented via performing the homology modeling, Sanger sequencing, and cell culture modifications. The other approach is the replicate study in an independent validation cohort.

Examples for the generation of clinical recommendations for the variants using in silico analysis of WGS PGx data were done before. The related studies showed the PGx dosage recommendations are heavily influenced by the higher availability of genotyping results, which may lead to more clinical evidence too [30]. Yet, the most important barrier to routine implementation of NGS technologies for PGx tests in clinical centers is the huge amount of uncertain and unknown significant variants in the results (group 5), which need to be confirmed and validated before considered as the influential elements in treatment decision and prescription modification. In addition to some basic problems in using NGS methods, such as poor coverage of the specific parts of the genome, false-positive results in short reads, ignoring many non-coding variants in targeted panels and WES, missing some homopolymer regions, pseudogenes, and GC rich, diverse efficiency for genome capturing due to the utilization of different kits and reagents, etc. [31], any novel or incidental markers still must go through the different validation steps, to be connected to drug-related phenotypes in patients. While looking for previous clinical reports and similar investigations, current approaches in dealing with PGx variants in group 5 are including the computational methods and in-vitro functional analysis of the variants. As the number of altered alleles could be high in NGS data, applying the computational analysis techniques and starting with categorizing, filtering, and functional annotating the variants across the RefSeq and other databases, such as dbSNP or dbNSFP, by special bioinformatics tools, such as *VAT*, *VarAFT*, *ANNOVAR*, etc., is inevitable. Then, the prediction of potentially damaging, deleterious, and/or functionally neutral non-synonymous variants will be performed via the algorithms as mentioned earlier. Currently, the mutual beliefs for PGx data analysis are the combined utilization of 6 to 7 of such prediction tools and choosing those variants, which are commonly introduced as pathogen/likely pathogen in all applied software, according to reliable reference guidelines, such as those given by ACMG, CAP, and CPIC [27]. While there is no universal and widely accepted functional prediction software package, the number of introduced PGx specific analysis tools, such as *Stargazer*, *Astrolabe*, *PharmCAT*, *PHARMIP*, etc., are increasing rapidly in a fast-developing mode. Hence, integrating them in applied algorithms seems necessary. Table 3 listed some of these special data mining and visualization tools, which are used or considered to be useful in PGx data management. We will talk about the limitations of common analysis facilities later in the discussion section. Next is the pathway mapping of the selected variants against the general and specialized free reference sources, such as PharmGKB, String-db, DAVID, KEGG, etc., to find out about the potential gene–drug and protein–protein interactions. Finally, allele frequency and population derived variant analysis could be achieved through comparing with comprehensive surveys (1000 genome, ExAc, HapMap, ESP, gnomAD, GME) [32]. Moreover, laboratory confirmative assays and characterization could be implemented for just top prioritized functional variants, to roll out any false-positive result and be assured of the real harmful effects on drug response. The final clinical assessments (if it's available) support the necessity for genotype–phenotype correlation procedures too.

Table 3. Special data mining and visualization tools and algorithms, used in PGx data analyzing and phenotype prediction.

Software	*Applications*	*Link*	*Reference*
SIFT	SIFT (Sorting Intolerant From Tolerant) is an online program that predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids.	https://sift.bii.a-star.edu.sg/	[33]
PolyPhen-2	PolyPhen-2 (Polymorphism Phenotyping v2) is a tool that predicts the possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations.	http://genetics.bwh.harvard.edu/pph2/	[34]
LOFTEE	Loss-Of-Function Transcript Effect Estimator is a tool to identify LoF (loss-of-function) effects of variations. LOFTEE also makes predictions of another splice (OS) variants that may cause LoF by disrupting normal splicing patterns.	http://www.atgu.mgh.harvard.edu/resources/software/	[35]
VAT	Variant Annotation Tool is a computational framework to functionally annotate variants in personal genomes using a cloud-computing environment.	http://vat.gersteinlab.org/	[36]
VarAFT	Variant Annotation and Filter Tool is for the identification of disease-causing mutations in human genetics. The software improves annotation and filtration steps.	https://varaft.eu/	[37]

Table 3. *Cont.*

Software	*Applications*	*Link*	*Reference*
EV mutation	An online free tool for predicting the mutation effects from sequences.	https://marks.hms.harvard.edu/evmutation/	[38]
UCSF chimera package	UCSF Chimera is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and animations can be generated. The Resource for Biocomputing, Visualization, and Informatics (RBVI) and its precursor, which is interactive software tools and advanced web-based computational resources that provide integrated visualizations and analyses of molecular structures and related non-structural biological information.	https://www.cgl.ucsf.edu/chimera/	[39]
ICM-Molsoft	ICM-Pro empowers a biologist or chemist by providing a high-quality protein structure analysis, modeling, and docking desktop software environment. Main features include: analyze sequences and alignments, inspect protein structure, study pockets, and bound ligands and drugs, create surfaces, calculate electrostatics, make mutations, predict ligand binding sites, predict protein–protein interaction sites, perform small molecule and protein–protein docking, and design ligands.	http://www.molsoft.com/icm_pro.html	-
EVfold	EVfold uses an evolutionary variation to calculate a set of co-evolved residue pairs in a protein family using a global approach called maximum entropy, formally similar to partial correlations.	http://evfold.org/evfold-web/evfold.do	[40,41]

Table 3. *Cont.*

Software	*Applications*	*Link*	*Reference*
xBrowse	xBrowse is a platform for studying rare genetic diseases. It was built to provide genetic researchers and clinical geneticists a collaborative way to search for the causes of genetic disease using exome sequencing data. xBrowse accepts as input a set of variant calls from a whole exome or whole genome sequencing study for further processing and annotation. Currently, the only accepted input format is a VCF file produced by the GATK pipeline.	http://www.atgu.mgh.harvard.edu/resources/software/	.
PLINK	PLINK/SEQ is an open-source C/C++ library for working with human genetic variation data. The specific focus is to provide a platform for analytic tool development for variation data from large-scale resequencing and genotyping projects, particularly whole-exome and whole-genome studies. It is independent of (but designed to be complementary to) the existing PLINK package.	https://atgu.mgh.harvard.edu/plinkseq/	[42]
SKAT	SKAT is a Single Nucleotide Polymorphism (SNP)-set (e.g., a gene or a region) level test for association between a set of rare (or common) variants and dichotomous or quantitative phenotypes, SKAT aggregates individual score test statistics of SNPs in a SNP set and efficiently computes SNP-set level *p*-values, e.g., a gene or a region-level *p*-value, while adjusting for covariates, such as principal components to account for population stratification. SKAT also allows for power/sample size calculations for designing sequence association studies.	www.hsph.harvard.edu/skat	[43]

Table 3. *Cont.*

Software	*Applications*	*Link*	*Reference*
Mutation Assessor	This server predicts the functional impact of amino-acid substitutions in proteins, such as mutations discovered in cancer or missense polymorphisms. The functional impact is assessed based on the evolutionary conservation of the affected amino acid in protein homologs.	http://mutationassessor.org/r3/	[44]
MutationTaster	MutationTaster is a free web-based application to evaluate DNA sequence variants for their disease-causing potential. The software performs a battery of in silico tests to estimate the impact of the variant on the gene product/protein.	http://www.mutationtaster.org/	[45]
PANTHER	The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System was designed to classify proteins (and their genes) to facilitate high-throughput analysis. PANTHER is defined as a method to predict the functional effect of missense variants based on sequence information.	http://www.pantherdb.org/	[46]
PhD-SNP	An SVM-based classifier for the prediction of variant pathogenicity according to sequence profiles.	http://snps.biofold.org/phd-snpg/	[47]
Varscan2	An analysis tool, for the detection of somatic mutations and copy number alterations (CNAs) in exome data from tumor–normal pairs.	http://varscan.sourceforge.net/	[48]
SPLINTER	Detects and quantifies short Insertion–Deletions (InDels) and substitutions in large pools. SPLINTER allows accurate detection and quantification of short insertions, deletions, and substitutions by integrating information from the synthetic DNA library to tune SPLINTER and quantify specificity and sensitivity for every experiment to accurately detect and quantify InDels and substitutions.	https://omictools.com/splinter-tool	[49]

Table 3. *Cont.*

Software	*Applications*	*Link*	*Reference*
GeneSplicer	GeneSplicer is a new, flexible system for detecting splice sites in the genomic DNA of various eukaryotes and predicting the variant effects on the related protein(s).	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml	[50]
NMD Classifier	NMD is a tool for systematic classification of nonsense-mediated decay events for either annotated or de novo assembled transcripts.	https://sourceforge.net/projects/transcriptome-analysis/files/NMD_Classifier.tar.gz	[51]
mrSNP	mrSNP provides a web service for researchers working especially with RNA-Seq Data, to predict the impact of an SNP in a 3UTR on miRNA binding.	https://tools4mirs.org/software/mirna_snp_analysis/mrsnp/	[52]
GenoCanyon	GenoCanyon is a whole-genome functional annotation approach based on unsupervised statistical learning. It integrates genomic conservation measures and biochemical annotation data to predict the functional potential at each nucleotide, both in coding, and non-coding regions.	http://genocanyon.med.yale.edu/	[53]
ANNOVAR	ANNOVAR is an efficient software tool to utilize up-to-date information to functionally annotate genetic variants detected from diverse genomes (including human genome hg18, hg19, hg38, as well as mouse, worm, fly, yeast, and many others). Given a list of variants with chromosome, start position, end position, reference nucleotide, and observed nucleotides, ANNOVAR can perform: - Gene-based annotation; - Region-based annotation; - Filter-based annotation, etc.	http://annovar.openbioinformatics.org/en/latest/	[54]

Table 3. *Cont.*

Software	*Applications*	*Link*	*Reference*
CADD	CADD is a tool for scoring the deleteriousness of single nucleotide variants as well as insertion/deletion variants in the human genome. It integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated mutations. C-scores strongly correlate with allelic diversity, the pathogenicity of both coding and non-coding variants.	https://cadd.gs.washington.edu/	[55,56]
Provean	Provean is a software tool that predicts whether an amino acid substitution or InDel has an impact on the biological function of a protein. It is useful for filtering sequence variants to identify non-synonymous or InDel variants that are predicted to be functionally important.	http://provean.jcvi.org/index.php	[57,58]
ESEfinder	ESEfinder is a web-based resource that facilitates rapid analysis of exon sequences to identify putative exonic splicing enhancers, responsive to the human SR proteins SF2/ASF, SC35, SRp40, and SRp55, and to predict whether exonic mutations disrupt such elements.	http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home	[59]
VarSeq	VarSeq is an intuitive, integrated software solution for tertiary analysis of next generation sequencing (NGS) data. With VarSeq workflows can be automated and analyzing variants for gene panels, exomes, and whole genomes is possible. Moreover, the tool shows the ability to integrate with new resources and databases for advanced and customized variant analysis.	https://www.goldenhelix.com/products/VarSeq/	[60]
FATHMM	A high-throughput web-server capable of predicting the functional consequences of both coding variants, i.e., non-synonymous single nucleotide variants (nsSNVs), and non-coding variants in the human genome.	http://fathmm.biocompute.org.uk/	[61]

Table 3. *Cont.*

Software	*Applications*	*Link*	*Reference*
GERP++	Genomic Evolutionary Rate Profiling (GERP) identifies constrained elements in multiple alignments by quantifying substitution deficits.	http://mendel.stanford.edu/SidowLab/downloads/gerp/	[62]
SiPhy	SiPhy implements rigorous statistical tests to detect bases under selection from multiple alignment data. It takes full advantage of deeply sequenced phylogenies to estimate either unlikely substitution patterns as well as slowdowns or accelerations in mutation rates.	http://portals.broadinstitute.org/genome_bio/siphy/index.html	-
Stargazer	Stargazer is a bioinformatics tool for calling star alleles (haplotypes) in PGx genes using data from NGS or SNP array. Stargazer can accept NGS data from both whole genome sequencing (WGS) and targeted sequencing. Stargazer identifies star alleles by detecting SNVs, InDels, and SVs. Stargazer can detect complex SVs including gene deletions, duplications, and hybrids by calculating paralog-specific copy numbers from read depth.	https://stargazer.gs.washington.edu/stargazerweb/	[63]
PharmCAT	A tool to extract all CPIC guideline gene variants from a genetic dataset (represented as a VCF file), interpret the variant alleles and generate a report.	https://github.com/PharmGKB/PharmCAT	[64,65]
PHARMIP	An in silico method to predict genetics that underpin adverse drug reactions. The tool can be used to reveal genetic risk factors for certain drug ADRs.	http://www.lilab-ecust.cn/pharmmapper/	[66]
PharmVar API	An online source for access to all or selected data of the Pharmacogene Variation Consortium (PharmVar) database.	https://www.pharmvar.org/documentation	-

Table 3. *Cont.*

Software	*Applications*	*Link*	*Reference*
Astrolabe	Astrolabe is software for the translation of whole genome sequence data into pharmacogenetic information that can be used to guide medication selection, dosing, and prescription. It was initially developed under the name Constellation for the *CYP2D6* gene, then extended to *CYP2C9* and *CYP2C19* with additional genes in the process of being validated. Astrolabe is integrated with the PharmVar database	https://childrensmercy.org/genomesoftwareportal/Software/Index/	[67]
Aldy	Aldy performs allelic decomposition of highly polymorphic, multi-copy genes by using whole or targeted genome sequencing data. For a large diverse sequencing data set, Aldy identifies multiple rare and novel alleles for several important pharmacogenes, significantly improving upon the accuracy and utility of current genotyping assays.	http://aldy.csail.mit.edu.	[68]
Cypripi	An algorithm to computationally infer *CYP2D6* genotype at base pair resolution from high throughput sequencing data. It can resolve complex genotypes, including alleles that are the products of duplication, deletion, and fusion events involving *CYP2D6* and its evolutionarily related cousin *CYP2D7*.	http://sfu-compbio.github.io/cypiripi/	[69]

5. Discussion

NGS technologies have been used in several PGx studies in recent years. Based on the employed platforms, the acquired data analyzed through different approaches. Due to the lower amount of identified variants (mostly known alleles), finding the FGVs and phenotype prediction is usually easier when targeted sequencing for a specific set of the gene (panels) is performed as the selected method. WES and WGS, however, show a lot of obstacles when applied for a PGx analysis and this is mainly because of the huge number of functionally unknown and unreported alterations in a patient's genetic profile [70]. Moreover, some intrinsic and substantial complications for PGx tests including the presence of germline mutations with necessary haplotype detection and phase definition in patients, going through specific pharmacogenes with a role in different sophisticated cellular pathways (i.e., *ACE*), following environmental and epigenetic modifications on drug-related genes, working with challenging and problematic variants, in particular drug-related genes (i.e., *CYP2D6* with close pseudogenes and many unknown and novel variants in diverse

populations, different functional tandem repeat variants in the non-coding part of *UGT1A1* gene, etc.), and most important of them the lack of previous knowledge on possible phenotype modifications for many genetic changes (as PGx is a pre-emptive genotyping test in numerous cases) can potentially increase the difficulties in variant analysis and pose the clear effects on changing the drug responses in individuals. Albeit, providing more genotype to phenotype translation methods by reference organizations and guideline developers will result in more consistent genotype interpretation in both clinical and research area [71].

Despite the challenges, the number of publications for NGS derived PGx data analysis are still significant. Gordon et al. successfully identified common, rare, and novel variants in 84 clinically actionable drug-related genes in more than 280 individuals through a targeted resequencing custom panel. They used deep coverage of the known genes to follow both previously recognized and possible novel variants. New potentially deleterious nonsense and missense variants across some VIPs were selected for more genotype-phenotype association studies to find any relation with particular traits (group1, 2, and 5 of the PGx variants). Moreover, actionable plus rare unreported variants in absorption, distribution, metabolism, and excretion (ADME) core genes revealed in 114 drug genes in 376 people by Han and colleagues. The number of variants in each gene (normalized based on gene length), MAF, and novelty assessed and compared to open genotyping datasets (group2, 4, and 5). In silico functional assessments performed by the prediction tools, such as *SIFT*, *PolyPhen2*, and *CAAD,* and deleterious rare-novel variants in some of VIPs evaluated by in-vitro analysis to find impaired functions evidence. Moreover, additional and novel faraway variants (group 5), contributed to the alteration of estrogen receptor binding site and breast cancer risk identified in 400 patients by NGS deep sequencing and functional genomics. As the number of investigated genes was low, any novel PGx variant was confirmed through the laboratory tests, such as chromatin immunoprecipitation (ChIP), gene expression analysis, and protein degradation assays [72–74]. Other utilizations also brought more unprecedented results for clinical PGx investigations. For example variants and haplotype detection of challenging ADME genes were successfully achieved in three core pharmacogenes (*CYP2D6*, *HLA-A,* and *HLA-B*) by applying the long read sequencers (group1 and 2). All the SNVs, CNVs, and InDels were revealed through the utilization of customized long-range PCR and the subsequent NGS machine (MinION nanopore sequencer) [75]. Moreover, 17,733 ADME variants per individual were detected in 231 genes. In addition to known PGx markers, the latter included 1012 novel variants with potential deleterious functions identified in exons, introns, gene promoters, and proximal regulatory regions. The authors reanalyzed WGS provided data to find different PGx markers in close to 500 individuals. In silico analysis used the ANNOVAR tool for annotation and dbSNP137 and Complete Genomics public server for novelty assessments. Functional assays were also predicted via SIFT and Provean algorithms (group1, 2, 4, and 5) [12]. In another effort, whole genome sequencing (WGS) in PGx analysis revealed 227 common and 466 rare population-specific potentially functional SNVs, including 74 novel variants in 437 drug genes (group1, 2, and 5). Variant analysis computational workflow consisted of ANNOVAR and dbSNP138 for variant annotation, *SIFT,* and *PolyPhen2* for functional effect analysis of novel non-synonymous coding SNVs, mapping the deleterious variants with PharmGKB and DrugBank, and finally *PLINK* and *VCFTools* for reaching allele frequencies and validation through 1000 genome and HapMap databases. In the end, a drug pathway map for functionally impaired pharmacogenes displayed, using identified deleterious variants [32]. Even the PGx-specific panel with high accuracy designed and identified clinically relevant variants in 39 genes including *CYP2D6* CNV and *UGT1A1*28* TAA repeats in promoter in addition to allele frequency and homozygosity in 235 patients. Common in-vitro and bioinformatics tools used for both known and novel variant detection rate accuracy and sensitivity (group1, 2, and 5) [76]. Finally, a comprehensive usage for NGS methods can be found in Price and his team effort, which applied exome sequencing for 21,000 human genes and revealed novel genetic loci with a strong association with on-treatment reactivity

and hereditability of platelet and clopidogrel response. Once again, novel loci and related variants in addition to known PGx markers were depicted by common data interpretation pipeline and proved the NGS methods as a powerful approach in unavailing PGx variants in clinical studies [77].

Two important points could be mentioned from the above investigations as well. As the majority of functional prediction tools and algorithms are relying on evolutionary conservation and therefore will not be completely fit with the pharmacogenes (poorly conserved) and show low predictive accuracy as the conventional algorithms (up to 50%), most of the studies emphasize combined utilization of such tools in in silico phenotype prediction for novel variants and introduced various software in each report. This may remind the necessity of the attitude for new PGx data in high throughput sequencing methods, as they are not observable in many cases (pre-emptive genotyping). Recent efforts, however, have been focused on developing new pharmacogene optimized frameworks with more relation to PGx data assessment through the integration of specific algorithms or presenting the allele dedicated for pharmacovariant calling and showed to be more compatible with ADME genes with a higher rate of sensitivity and specificity (90–99%) [27,63]. Other PGx specialized projects are also recently developed a pharmacogenomics clinical annotation tool (*PharmCAT*) and tried to reveal which patients in a clinical dataset include the variants of interest [65].

The second point is the ability of NGS technologies to the detection of any kinds of PGx variants in clinical practice. They have introduced several novel PGx markers successfully and the fact may indicate the faster incorporation of PGx test results into the future precision medicine as well. However, there are still essential issues with high importance in the field, which need to be addressed properly. For example, if the particular novel variant causes a loss of function or gain of function effects on the related protein(s) (making a poor or rapid metabolizer) in tested individuals and also possible misinterpreting of VUS in the result, which may lead to ignore or miss the functional variants in pharmacogenes. Such complexities must be followed by the in-vitro assessments in addition to appropriate pre and post-test counseling for individuals [28,78].

The intricacies are not limited to the detection of variants, but the nature of drug actions according to particular alleles too. Investigations displayed the dual or multiple impacts of some specific pharmacovariants toward the different diseases and/or drugs (Table 4). Furthermore, a certain drug could be the substrate for more than one P450 family and metabolized by different enzymes (i.e., CYP1A2, CYP2C19, and CYP2D6 for antidepressant amitriptyline) [79]. Such scenarios complicate the true functional assessment of pharmacovariants, especially in high throughput sequencing data. Because of that, a comprehensive literature search, replicate studies, and wet lab analysis of the newly identified genetic markers in drug-related genes must be taken into account before any prescription considerations in the clinical setting.

Table 4. Examples of different outcomes for one particular allele/diplotype of *CYP2D6* in different disorders and drugs.

Disease/Disorder	*Drug*	*Gene*	*Diplotype or Allele*	*Decreased Response*	*Increased Response*	*Low Plasma Concentration*	*High Plasma Concentration*	*Toxicity*	*Level of Evidence*	*Reference*
Depressive Disorder Mental Disorders	Paroxetine	*CYP2D6*	*1/*1xN #	-	✓	✓	-	-	1A	[80]
Nausea and Vomiting after Chemotherapy	Ondansetron	*CYP2D6*	*1/*1xN	✓	-	-	✓	?	1A	[81]
Mental Disorders	Desipramine	*CYP2D6*	*1xN	✓	-	-	✓	?	2A	[82]
Alzheimer Disease	Donepezil	*CYP2D6*	*1/*1xN	-	✓	✓	-	-	3	[83]
Pain	Codeine	*CYP2D6*	*1/*1xN	-	✓	✓	-	✓	1A	[84,85]

Gene duplication, which resulted in ultra-rapid metabolizer. * is a standardized nomenclature system used for various haplotypes and alleles in Cytochrome P450 family pharmacogenes. The level of evidence is adopted from PharmGKB [19]. ✓: *Yes*, *?: unknown*, *-: not applicable.*

6. Conclusions

The field of pharmacogenomics faces several challenges throughout the process of the identification of pharmacogenomic variants and their implementation in clinical practices. Many of these challenges arise at the genomics level, including the statistical considerations associated with the design of the clinical trial and genome-wide association studies (GWAS), a large number of candidate variants compared to available samples ($p > n$), the lack of reproducibility in independent studies and determining the functional impact of variants on drug response. In the age of PGx and personalized drug therapy, using the high throughput sequencing approaches will assist the translation of different pharmacovariants into clinical care. As mentioned before, for moving genomic medicine toward personalized drug therapy, there should be a genetic screening test, which fits all ethnicities [12]. NGS, as a time and cost-effective and highly accurate genotyping method, shows the huge benefits for patients PGx clinical assessments. Hence, it would be highly possible for the investigators and clinicians to encounter new and rare population-specific variants during a PGx test. To deal with different NGS derived PGx variants in clinics, all healthcare professionals need to know the classification and interpretation algorithms for such markers properly.

Author Contributions: A.T. provided the idea, designed the study, conducted the literature search, and wrote the entire manuscript. N.W.-K. and R.P. searched for the literature and edited the manuscript. W.M. supervised the study thoroughly and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This article was conducted within the projects, which received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 754432, and the Polish Ministry of Science and Higher Education, from financial resources for science in 2018–2023 granted for the implementation of an international co-financed project.

Conflicts of Interest: The authors declare that there is no conflict of interest during the preparation of this manuscript.

Abbreviations

ACMG	American College of Medical Genetics and Genomics
ADME	Absorption, Distribution, Metabolism, and Excretion
CADD	Combined Annotation-Dependent Depletion
CAP	College of American Pathologists
CNV	Copy Number Variation
CPIC	The Clinical Pharmacogenetics Implementation Consortium
DPWG	Dutch Pharmacogenetics Working Group
FDA	Food and Drug Administration
FGV	Functional Genetic Variation
GWAS	Genome-Wide Association Studies
IF	Incidental Findings
InDel	Insertion–Deletion
MAF	Minor Allele Frequency
NCI	National Cancer Institute
NGS	Next Generation Sequencing
PDG	Pharmacogenomics Dosage Guidelines
PGRN	The Pharmacogenomics Research Network
PGx	Pharmacogenomics
PharmCAT	Pharmacogenomics Clinical Annotation Tool

PharmGKB	Pharmacogenomics Knowledge Base
Provean	Protein Variation Effect Analyzer
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
SV	Structural Variants
VAT	Variant Annotation Tool
VarAFT	Variant Annotation and Filter Tool
VCF	Variant Call Format
VIP	Very Important Pharmacogene
VUS	Variants with Unknown clinical Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

References

1. Sim, S.; Kacevska, M.; Ingelman-Sundberg, M. Pharmacogenomics of drug-metabolizing enzymes: A recent update on clinical implications and endogenous effects. *Pharm. J.* **2013**, *13*, 1–11. [CrossRef] [PubMed]
2. Kozyra, M.; Ingelman-sundberg, M.; Lauschke, V.M. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet. Med.* **2017**, *19*, 20–29. [CrossRef] [PubMed]
3. Schärfe, C.P.I.; Tremmel, R.; Schwab, M.; Kohlbacher, O.; Marks, D.S. Genetic variation in human drug-related genes. *Genome Med.* **2017**, *9*, 117. [CrossRef]
4. 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. [CrossRef] [PubMed]
5. Walczak, M.; Skrzypczak-Zielinska, M.; Plucinska, M.; Zakerska-Banaszak, O.; Marszalek, D.; Lykowska-Szuber, L.; Stawczyk-Eder, K.; Dobrowolska, A.; Slomski, R. Long-range PCR libraries and next generation sequencing for pharmacogenetic studies of patients treated with anti-TNF drugs. *Pharm. J.* **2019**, *19*, 358–367. [CrossRef]
6. Santos, M.; Niemi, M.; Hiratsuka, M.; Kumondai, M.; Ingelman-Sundberg, M.; Lauschke, V.M.; Rodríguez-Antona, C. Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genet. Med.* **2018**, *20*, 622–629. [CrossRef]
7. Cousin, M.A.; Matey, E.T.; Blackburn, P.R.; Boczek, N.J.; McAllister, T.M.; Kruisselbrink, T.M.; Babovic-Vuksanovic, D.; Lazaridis, K.N.; Klee, E.W. Pharmacogenomic findings from clinical whole exome sequencing of diagnostic odyssey patients. *Mol. Genet. Genome Med.* **2017**, *5*, 269–279. [CrossRef]
8. Klein, K.; Tremmel, R.; Winter, S.; Fehr, S.; Battke, F.; Scheurenbrand, T.; Schaeffeler, E.; Biskup, S.; Schwab, M.; Zanger, U.M. A New Panel-Based Next Generation Sequencing Method for ADME Genes Reveals Novel Associations of Common and Rare Variants With Expression in a Human Liver Cohort. *Front. Genet.* **2019**, *10*, 7. [CrossRef]
9. Hovelson, D.H.; Xue, Z.; Zawistowski, M.; Ehm, M.G.; Harris, E.C.; Stocker, S.L.; Gross, A.S.; Jang, I.-J.; Ieiri, I.; Lee, J.-E.; et al. Characterization of ADME gene variation in 21 populations by exome sequencing. *Pharm. Genome* **2017**, *27*, 89. [CrossRef]
10. Ortega, V.E.; Meyers, D.A. Pharmacogenetics: Implications of race and ethnicity on defining genetic profiles for personalized medicine. *J. Allergy Clin. Immunol.* **2014**, *133*, 16–26. [CrossRef]
11. Ahn, E.; Park, T. Analysis of population-specific pharmacogenomic variants using next generation sequencing data. *Sci. Rep.* **2017**, *7*, 8416. [CrossRef] [PubMed]
12. Mizzi, C.; Peters, B.; Mitropoulou, C.; Mitropoulos, K.; Katsila, T.; Agarwal, M.R.; Van Schaik, R.H.; Drmanac, R.; Borg, J.; Patrinos, G.P. Personalized pharmacogenomics profiling using whole-genome sequencing. *Pharmacogenomics* **2014**, *15*, 1223–1234. [CrossRef] [PubMed]
13. Lalioti, M.; Gerasimova, T.; Zattas, D.; Anastasakis, D.; Seli, E.; Sakkas, D. *A Deleted Form of FSH Receptor, Found in Women Undergoing Infertility Treatment, Impairs the Function of the Normal Receptor When Co-Expressed In Vitro*; Oxford University Press: Oxford, UK, 2010.
14. Hoffman, J.M.; Dunnenberger, H.M.; Hicks, J.K.; Caudle, K.E.; Carrillo, M.W.; Freimuth, R.R.; Williams, M.S.; Klein, T.E.; Peterson, J.F. Developing knowledge resources to support precision medicine: Principles from the Clinical Pharmacogenetics Implementation Consortium (CPIC). *J. Am. Med Inform. Assoc.* **2016**, *23*, 796–801. [CrossRef] [PubMed]
15. Relling, M.; Klein, T. CPIC: Clinical pharmacogenetics implementation consortium of the pharmacogenomics research network. *Clin. Pharmacol. Ther.* **2011**, *89*, 464–467. [CrossRef] [PubMed]
16. Barbarino, J.M.; Whirl-Carrillo, M.; Altman, R.B.; Klein, T.E. PharmGKB: A worldwide resource for pharmacogenomic information. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2018**, *10*, e1417. [CrossRef]
17. Tilleman, L.; Weymaere, J.; Heindryckx, B.; Deforce, D.; Van Nieuwerburgh, F. Contemporary pharmacogenetic assays in view of the PharmGKB database. *Pharmacogenomics* **2019**, *20*, 261–272. [CrossRef]
18. Swen, J.J.; Nijenhuis, M.; De Boer, A.; Grandia, L.; Der Zee, A.M.-V.; Mulder, H.; Rongen, G.; Van Schaik, R.; Schalekamp, T.; Touw, D.; et al. Pharmacogenetics: From Bench to Byte—An Update of Guidelines. *Clin. Pharmacol. Ther.* **2011**, *89*, 662–673. [CrossRef]
19. PharmGKB. Available online: https://www.pharmgkb.org/page/clinAnnLevels (accessed on 14 April 2020).

20. CPIC. Available online: https://cpicpgx.org/genes-drugs (accessed on 14 April 2020).
21. DPWG. Available online: https://www.pharmgkb.org/page/dpwg (accessed on 14 April 2020).
22. Caudle, K.E.; Gammal, R.S.; Whirl-Carrillo, M.; Hoffman, J.M.; Relling, M.V.; Klein, T.E. Evidence and resources to implement pharmacogenetic knowledge for precision medicine. *Am. J. Health Syst. Pharm.* **2016**, *73*, 1977–1985. [CrossRef]
23. Relling, M.V.; Evans, W.E. Pharmacogenomics in the clinic. *Nature* **2015**, *526*, 343–350. [CrossRef]
24. Whirl-Carrillo, M.; McDonagh, E.M.; Hebert, J.; Gong, L.; Sangkuhl, K.; Thorn, C.; Altman, R.B.; Klein, T.E. Pharmacogenomics Knowledge for Personalized Medicine. *Clin. Pharmacol. Ther.* **2012**, *92*, 414–417. [CrossRef]
25. Aslibekyan, S.; Claas, S.A.; Arnett, D.K. To replicate or not to replicate: The case of pharmacogenetic studies: Establishing validity of pharmacogenomic findings: From replication to triangulation. *Circ. Cardiovasc. Genet.* **2013**, *6*, 409–412. [CrossRef] [PubMed]
26. Lee, E.M.; Xu, K.; Mosbrook, E.; Links, A.; Guzman, J.; Adams, D.R.; Flynn, E.; Valkanas, E.; Toro, C.; Tifft, C.J.; et al. Pharmacogenomic incidental findings in 308 families: The NIH Undiagnosed Diseases Program experience. *Genet. Med.* **2016**, *18*, 1303–1307. [CrossRef] [PubMed]
27. Zhou, Y.; Mkrtchian, S.; Kumondai, M.; Hiratsuka, M.; Lauschke, V.M. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharm. J.* **2019**, *19*, 115–126. [CrossRef] [PubMed]
28. Devarajan, S.; Moon, I.; Ho, M.-F.; Larson, N.B.; Neavin, D.R.; Moyer, A.M.; Black, J.L.; Bielinski, S.J.; Scherer, S.E.; Wang, L.; et al. Pharmacogenomic Next-Generation DNA Sequencing: Lessons from the Identification and Functional Characterization of Variants of Unknown Significance in CYP2C9 and CYP2C19. *Drug Metab. Dispos.* **2019**, *47*, 425–435. [CrossRef]
29. Twesigomwe, D.; Wright, G.E.; Drögemöller, B.I.; da Rocha, J.; Lombard, Z.; Hazelhurst, S. A systematic comparison of pharmacogene star allele calling bioinformatics algorithms: A focus on CYP2D6 genotyping. *NPJ Genom. Med.* **2020**, *5*, 1–11. [CrossRef]
30. Huser, V.; Cimino, J.J. Providing pharmacogenomics clinical decision support using whole genome sequencing data as input. *AMIA Jt. Summits Transl. Sci. Proc.* **2013**, *2013*, 81.
31. Shigemizu, D.; Fujimoto, A.; Akiyama, S.; Abe, T.; Nakano, K.; Boroevich, K.A.; Yamamoto, Y.; Furuta, M.; Kubo, M.; Nakagawa, H.; et al. A practical method to detect SNVs and indels from whole genome and exome sequencing data. *Sci. Rep.* **2013**, *3*, 2161. [CrossRef]
32. Sivadas, A.; Salleh, M.; Teh, L.; Scaria, V. Genetic epidemiology of pharmacogenetic variants in South East Asian Malays using whole-genome sequences. *Pharm. J.* **2017**, *17*, 461–470. [CrossRef]
33. Pauline, C.N.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [CrossRef]
34. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **2013**, *76*, 7–20. [CrossRef]
35. MacArthur, D.G.; Balasubramanian, S.; Frankish, A.; Huang, N.; Morris, J.; Walter, K.; Jostins, L.; Habegger, L.; Pickrell, J.K.; Montgomery, S.B.; et al. A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. *Science* **2012**, *335*, 823–828. [CrossRef] [PubMed]
36. Habegger, L.; Balasubramanian, S.; Chen, D.Z.; Khurana, E.; Sboner, A.; Harmanci, A.; Rozowsky, J.; Clarke, D.; Snyder, M.; Gerstein, M. VAT: A computational framework to functionally annotate variants in personal genomes within a cloud-computing environment. *Bioinformatics* **2012**, *28*, 2267–2269. [CrossRef] [PubMed]
37. Desvignes, J.-P.; Bartoli, M.; Delague, V.; Krahn, M.; Miltgen, M.; Béroud, C.; Salgado, D. VarAFT: A variant annotation and filtration system for human next generation sequencing data. *Nucleic Acids Res.* **2018**, *46*, W545–W553. [CrossRef] [PubMed]
38. Hopf, T.A.; Ingraham, J.B.; Poelwijk, F.J.; Schärfe, C.P.; Springer, M.; Sander, C.; Marks, D.S. Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **2017**, *35*, 128–135. [CrossRef]
39. Huang, C.C.; Meng, E.C.; Morris, J.H.; Pettersen, E.F.; Ferrin, T.E. Enhancing UCSF Chimera through web services. *Nucleic Acids Res.* **2014**, *42*, W478–W484. [CrossRef]
40. Marks, D.S.; Colwell, L.J.; Sheridan, R.; Hopf, T.A.; Pagnani, A.; Zecchina, R.; Sander, C. Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS ONE* **2011**, *6*, e28766. [CrossRef]
41. Marks, D.S.; Hopf, T.A.; Sander, C. Protein structure prediction from sequence variation. *Nat. Biotechnol.* **2012**, *30*, 1072–1080. [CrossRef]
42. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]
43. Ionita-Laza, I.; Lee, S.; Makarov, V.; Buxbaum, J.D.; Lin, X. Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants. *Am. J. Hum. Genet.* **2013**, *92*, 841–853. [CrossRef]
44. Li, H. Tabix: Fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **2011**, *27*, 718–719. [CrossRef]
45. Schwarz, J.M.; Rödelsperger, C.; Schuelke, M.; Seelow, D. Mutation Taster evaluates diseasecausing potential of sequence alterations. *Nat. Methods* **2010**, *7*, 575–576. [CrossRef] [PubMed]
46. Thomas, P.D.; Muruganujan, A.; Ebert, D.; Huang, X.; Thomas, P.D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **2018**, *7*, D419–D426. [CrossRef]
47. Capriotti, E.; Fariselli, P. PhD-SNPg: A webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res.* **2017**, *45*, W247–W252. [CrossRef] [PubMed]

48. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **2012**, *22*, 568–576. [CrossRef]
49. Vallania, F.; Ramos, E.; Cresci, S.; Mitra, R.D.; Druley, T.E. Detection of Rare Genomic Variants from Pooled Sequencing Using SPLINTER. *J. Vis. Exp.* **2012**, *64*, e3943. [CrossRef] [PubMed]
50. Pertea, M.; Lin, X.; Salzberg, S.L. GeneSplicer: A new computational method for splice site prediction. *Nucleic Acids Res.* **2001**, *29*, 1185–1190. [CrossRef]
51. Hsu, M.-K.; Lin, H.-Y.; Chen, F.-C. NMD Classifier: A reliable and systematic classification tool for nonsense-mediated decay events. *PLoS ONE* **2017**, *12*, e0174798. [CrossRef]
52. Deveci, M.; Çatalyürek, Ü.V.; Toland, A.E. mrSNP: Software to detect SNP effects on microRNA binding. *BMC Bioinform.* **2014**, *15*, 73. [CrossRef]
53. Lu, Q.; Hu, Y.; Sun, J.; Cheng, Y.; Cheung, K.-H.; Zhao, H. A Statistical Framework to Predict Functional Non-Coding Regions in the Human Genome through Integrated Analysis of Annotation Data. *Sci. Rep.* **2015**, *5*, 10576. [CrossRef]
54. Yang, H.; Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* **2015**, *10*, 1556–1566. [CrossRef]
55. Kircher, M.; Witten, D.M.; Jain, P.; O'Roak, B.J.; Cooper, G.M.; Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **2014**, *46*, 310. [CrossRef] [PubMed]
56. Rentzsch, P.; Witten, D.; Cooper, G.M.; Shendure, J.; Kircher, M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **2019**, *47*, D886–D894. [CrossRef] [PubMed]
57. Choi, Y.; Chan, A.P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745–2747. [CrossRef] [PubMed]
58. Choi, Y.; Sims, G.E.; Murphy, S.; Miller, J.R.; Chan, A.P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE* **2012**, *7*, e46688. [CrossRef]
59. Cartegni, L.; Wang, J.; Zhu, Z.; Zhang, M.Q.; Krainer, A.R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **2003**, *31*, 3568–3571. [CrossRef]
60. VarSeq. VarSeq Brochures. Available online: https://www.goldenhelix.com/products/VarSeq/ (accessed on 12 April 2020).
61. Rogers, M.F.; Shihab, H.A.; Mort, M.; Cooper, D.N.; Gaunt, T.R.; Campbell, C. FATHMM-XF: Accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **2018**, *34*, 511–513. [CrossRef]
62. Davydov, E.V.; Goode, D.L.; Sirota, M.; Cooper, G.M.; Sidow, A.; Batzoglou, S. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLoS Comput. Biol.* **2010**, *6*, e1001025. [CrossRef]
63. Lee, S.-B.; Wheeler, M.M.; Patterson, K.; McGee, S.; Dalton, R.; Woodahl, E.L.; Gaedigk, A.; Thummel, K.E.; Nickerson, D.A. Stargazer: A software tool for calling star alleles from next generation sequencing data using CYP2D6 as a model. *Genet. Med.* **2019**, *21*, 361–372. [CrossRef]
64. Klein, T.E.; Ritchie, M.D. PharmCAT: A Pharmacogenomics Clinical Annotation Tool. *Clin. Pharmacol. Ther.* **2018**, *104*, 19–22. [CrossRef]
65. Sangkuhl, K.; Whirl-Carrillo, M.; Whaley, R.M.; Woon, M.; Lavertu, A.; Altman, R.B.; Carter, L.; Verma, A.; Ritchie, M.D.; Klein, T.E. Pharmacogenomics Clinical Annotation Tool (Pharm CAT). *Clin. Pharmacol. Ther.* **2020**, *107*, 203–210. [CrossRef]
66. Zidan, A.M.; Saad, E.A.; Ibrahim, N.E.; Mahmoud, A.; Hashem, M.H.; Hemeida, A.A. PHARMIP: An insilico method to predict genetics that underpin adverse drug reactions. *MethodsX* **2020**, *7*, 100775. [CrossRef] [PubMed]
67. Twist, G.P.; Gaedigk, A.; Miller, N.A.; Farrow, E.G.; Willig, L.K.; Dinwiddie, D.L.; Petrikin, J.; Soden, S.E.; Herd, S.; Gibson, M.; et al. Constellation: A tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene, CYP2D6, from whole-genome sequences. *NPJ Genome Med.* **2016**, *1*, 1–10. [CrossRef] [PubMed]
68. Numanagić, I.; Malikić, S.; Ford, M.; Qin, X.; Toji, L.; Radovich, M.; Skaar, T.C.; Pratt, V.M.; Berger, B.; Scherer, S.; et al. Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. *Nat. Commun.* **2018**, *9*, 1–11. [CrossRef]
69. Numanagić, I.; Malikić, S.; Pratt, V.M.; Skaar, T.C.; Flockhart, D.A.; Sahinalp, S.C. Cypiripi: Exact genotyping of CYP2D6 using high-throughput sequencing data. *Bioinformatics* **2015**, *31*, i27–i34. [CrossRef] [PubMed]
70. Katsila, T.; Patrinos, G.P. Whole genome sequencing in pharmacogenomics. *Front. Pharmacol.* **2015**, *6*, 61. [CrossRef]
71. Caudle, K.E.; Sangkuhl, K.; Whirl-Carrillo, M.; Swen, J.J.; Haidar, C.E.; Klein, T.E.; Gammal, R.S.; Relling, M.V.; Scott, S.A.; Hertz, D.L.; et al. Standardizing CYP 2D6 Genotype to Phenotype Translation: Consensus Recommendations from the Clinical Pharmacogenetics Implementation Consortium and Dutch Pharmacogenetics Working Group. *Clin. Transl. Sci.* **2020**, *13*, 116–124. [CrossRef]
72. Gordon, A.; Fulton, R.S.; Qin, X.; Mardis, E.R.; Nickerson, D.A.; Scherer, S. PGRNseq: A targeted capture sequencing panel for pharmacogenetic research and implementation. *Pharm. Genome* **2016**, *26*, 161. [CrossRef]
73. Han, S.; Park, J.; Lee, J.; Lee, S.; Kim, H.; Han, H.; Kim, Y.; Yi, S.; Cho, J.Y.; Jang, I.J.; et al. Targeted Next-Generation Sequencing for Comprehensive Genetic Profiling of Pharmacogenes. *Clin. Pharmacol. Ther.* **2017**, *101*, 396–405. [CrossRef]
74. Liu, D.; Ho, M.-F.; Schaid, D.J.; Scherer, S.E.; Kalari, K.; Liu, M.; Biernacka, J.; Yee, V.; Evans, J.; Carlson, E.; et al. Breast cancer chemoprevention pharmacogenomics: Deep sequencing and functional genomics of the ZNF423 and CTSO genes. *NPJ Breast Cancer* **2017**, *3*, 1–9. [CrossRef]

75. Ammar, R.; Paton, T.A.; Torti, D.; Shlien, A.; Bader, G.D. Long read nanopore sequencing for detection of HLA and CYP2D6 variants and haplotypes. *F1000Research* **2015**, *4*, 17. [CrossRef]
76. Gulilat, M.; Lamb, T.; Teft, W.A.; Wang, J.; Dron, J.S.; Robinson, J.F.; Tirona, R.G.; Hegele, R.A.; Kim, R.B.; Schwarz, U.I. Targeted next generation sequencing as a tool for precision medicine. *BMC Med. Genome* **2019**, *12*, 81. [CrossRef] [PubMed]
77. Price, M.J.; Carson, A.R.; Murray, S.S.; Phillips, T.; Janel, L.; Tisch, R.; Topol, E.; Levy, S. First pharmacogenomic analysis using whole exome sequencing to identify novel genetic determinants of clopidogrel response variability: Results of the genotype information and functional testing (gift) exome study. *J. Am. Coll. Cardiol.* **2012**, *59*, E9. [CrossRef]
78. Hoehe, M.R.; Timmermann, B.; Lehrach, H. Human Inter-Individual DNA Sequence Variation in Candidate Genes, Drug Targets, the Importance of Haplotypes and Pharmacogenomics. *Curr. Pharm. Biotechnol.* **2003**, *4*, 351–378. [CrossRef] [PubMed]
79. Strachan, T.; Goodship, J.; Chinnery, P. *Genetics and Genomics in Medicine*; Taylor & Francis: Abingdon, UK, 2014.
80. Gex-Fabry, M.; Eap, C.B.; Oneda, B.; Gervasoni, N.; Aubry, J.-M.; Bondolfi, G.; Bertschy, G. CYP2D6 and ABCB1 Genetic Variability: Influence on Paroxetine Plasma Level and Therapeutic Response. *Ther. Drug Monit.* **2008**, *30*, 474–482. [CrossRef] [PubMed]
81. Kaiser, R.; Sezer, O.; Papies, A.; Bauer, S.; Schelenz, C.; Tremblay, P.-B.; Possinger, K.; Roots, I.; Brockmöller, J. Patient-Tailored Antiemetic Treatment With 5-Hydroxytryptamine Type 3 Receptor Antagonists According to Cytochrome P-450 2D6 Genotypes. *J. Clin. Oncol.* **2002**, *20*, 2805–2811. [CrossRef] [PubMed]
82. Kawanishi, C.; Lundgren, S.; Ågren, H.; Bertilsson, L. Increased incidence of CYP2D6 gene duplication in patients with persistent mood disorders: Ultrarapid metabolism of antidepressants as a cause of nonresponse. A pilot study. *Eur. J. Clin. Pharmacol.* **2004**, *59*, 803–807. [CrossRef] [PubMed]
83. Noetzli, M.; Guidi, M.; Ebbing, K.; Eyer, S.; Wilhelm, L.; Michon, A.; Thomazic, V.; Stancu, I.; Alnawaqil, A.M.; Bula, C. Population pharmacokinetic approach to evaluate the effect of CYP2D6, CYP3A, ABCB1, POR and NR1I2 genotypes on donepezil clearance. *Br. J. Clin. Pharmacol.* **2014**, *78*, 135–144. [CrossRef]
84. Ciszkowski, C.; Madadi, P.; Phillips, M.S.; Lauwers, A.E.; Koren, G. Codeine, Ultrarapid-Metabolism Genotype, and Postoperative Death. *N. Engl. J. Med.* **2009**, *361*, 827–828. [CrossRef]
85. Gasche, Y.; Daali, Y.; Fathi, M.; Chiappe, A.; Cottini, S.; Dayer, P.; Desmeules, J. Codeine Intoxication Associated with Ultrarapid CYP2D6 Metabolism. *N. Engl. J. Med.* **2004**, *351*, 2827–2831. [CrossRef]

Research Article

Development of an extensive workflow for comprehensive clinical Pharmacogenomic profiling: *lessons from a pilot study on 100 Whole Exome Sequencing data*

Alireza Tafazoli, Maaike van der Lee, Jesse J. Swen, Anna Zeller, Natalia Wawrusiewicz-Kurylonek, Hailiang Mei, Ruben H.P. Vorderman, Krzysztof Konopko, Andrzej Zankiewicz, and Wojciech Miltyk*

The Pharmacogenomics Journal (Nature publishing group)

DOI: 10.1038/s41397-022-00286-4

IF: 3.550
MEiN: 140

The Pharmacogenomics Journal www.nature.com/tpj

ARTICLE OPEN

Check for updates

Development of an extensive workflow for comprehensive clinical pharmacogenomic profiling: lessons from a pilot study on 100 whole exome sequencing data

Alireza Tafazoli[1,2], Maaike van der Lee[3], Jesse J. Swen[3], Anna Zeller[2], Natalia Wawrusiewicz-Kurylonek[4], Hailiang Mei[5], Ruben H. P. Vorderman[5], Krzysztof Konopko[6], Andrzej Zankiewicz[6] and Wojciech Miltyk[1]✉

This pilot study is aimed at implementing an approach for comprehensive clinical pharmacogenomics (PGx) profiling. Fifty patients with cardiovascular diseases and 50 healthy individuals underwent whole-exome sequencing. Data on 1800 PGx genes were extracted and analyzed through deep filtration separately. Theoretical drug induced phenoconversion was assessed for the patients, using *sequence2script*. In total, 4539 rare variants (including 115 damaging non-synonymous) were identified. Four publicly available PGx bioinformatics algorithms to assign PGx haplotypes were applied to nine selected very important pharmacogenes (VIP) and revealed a 45–70% concordance rate. To ensure availability of the results at point-of-care, actionable variants were stored in a web-hosted database and PGx-cards were developed for quick access and handed to the study subjects. While a comprehensive clinical PGx profile could be successfully extracted from WES data, available tools to interpret these data demonstrated inconsistencies that complicate clinical application.

The Pharmacogenomics Journal; https://doi.org/10.1038/s41397-022-00286-4

INTRODUCTION

Pharmacogenomics (PGx) is aimed at reducing adverse drug reactions (ADRs) and lack of efficacy by adjusting drug therapy based on an individual's genetic profile. Many single gene-drug interactions have been described so far. For interactions with the highest evidence, guidelines and recommendations are available [1]. Not only patients, but also healthy individuals may benefit from PGx testing for future prescriptions by saving the PGx data in electronic health records (EHR) [2]. Currently, targeted genotyping is standard practice in most PGx laboratories for the identification of important variants in the pharmacogenes. However, these panel-based tests are not able to identify rare genomic variants which are expected to have a substantial impact on a patient's drug response. Sequencing-based methods, on the other hand, are capable of detecting most of the rare variants [3, 4]. These additional variants may help to better explain and predict drug-related phenotypes. Several groups have investigated the utility of next-generation sequencing (NGS) for PGx, both with the use of whole-exome sequencing (WES) as well as whole-genome sequencing (WGS) [5–11]. Such studies for example, demonstrated the utilization of WGS for the identification of putatively functional variants within well-known pharmacogenes. The result successfully represented the missing causative variants underlying drug response phenotypes [12]. However, state-of-the-art high throughput sequencing approaches result in a large amount of data, making it necessary to develop more powerful PGx-bioinformatics tools as well as assess the clinical validity and utility of sequencing-based tests [13]. Multiple tools have been developed and tested in NGS-based PGx studies [14, 15]. We provided a comprehensive review of such tools and their functional algorithms previously [16]. The performance of available haplotyping tools was also compared for *CYP2D6* before. The study showed that while the overall performance was good, there were discrepancies between the individual tools. Nevertheless, a comparison of the utility of these tools for clinical PGx samples and a wide range of genes is yet to be made [17]. In this pilot study, we aim to develop an approach for comprehensive clinical PGx profiling of 100 participants. Also, we introduce a method of deep filtration for dealing with variants in less-studied drug-related genes.

METHODS

Sample collection

Blood samples from 100 participants of a local and longitudinal observational biomedical project were obtained (50 cardiovascular patients with pulmonary hypertension and ischemic disease and 50 healthy individuals with common demographic features as the control group). The project was approved by the Medical University of Bialystok bioethics committee (approval code: R-I-002/630/2018) and all participants provided informed consent.

[1]Department of Analysis and Bioanalysis of Medicines, Faculty of Pharmacy with the Division of Laboratory Medicine, Medical University of Bialystok, 15-089 Bialystok, Poland. [2]Clinical Research Centre, University Hospital of Bialystok, 15-276 Bialystok, Poland. [3]Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands. [4]Department of Clinical Genetics, Medical University of Białystok, 15-089 Białystok, Poland. [5]Sequencing Analysis Support Core, Department of Biomedical Data Sciences, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands. [6]Department of Photonics, Electronics, and Lighting Technology, Faculty of Electrical Engineering, Bialystok University of Technology, 15-351 Bialystok, Poland. ✉email: Wojciech.miltyk@umb.edu.pl

Received: 3 July 2022 Revised: 20 July 2022 Accepted: 25 July 2022
Published online: 13 August 2022

Whole-exome sequencing and primary analysis of data

DNA extraction, NGS library preparation, and quality assessment were performed according to standard manufacturer protocols (Supplementary Material). Pre-Capture Pooling Human All Exon V7" was used. The SureSelect^{XT} kits provide a target enrichment system for Illumina paired-end multiplexed sequencing library preparation. Sequencing was performed using the Illumina NovaSeq 6000 instrument. A standard bioinformatics analysis pipeline for raw data was employed for both GRCh37 and GRCh38 genome builds (Supplementary Material). In short, low-quality bases (read depth <10) were omitted and reads were aligned to the reference genome with the BWA-mem (Burrows-Wheeler) algorithm.

Data filtration and functional assessment

For variant interpretation, annotation, and initial functional assessment *SnpEff* and *ENSEMBL-VEP* were used [18, 19]. For the PGx assessment, a list of 1800 drug-related genes was prepared. These genes include metabolizer enzymes, drug transporters, drug receptors, and drug-target molecules. They were collected from the *PharmGKB* comprehensive gene list (only genes with at least one annotated variant extracted) ($n = 1707$), all *CPIC* gene-drug lists ($n = 119$), and the FDA table of "Pharmacogenomic Biomarkers" in drug labeling ($n = 132$), plus a systematic search in PubMed for any unannotated but newly introduced drug-related genes ($n = 17$). The relevant keywords selected and specified period applied for choosing state-of-the-art articles. Both the abstract and main text were evaluated systematically and the final result was added to our comprehensive gene list (Supplementary Material). Duplicates were then removed and variant call format (VCF) files were filtered to contain only these 1800 genes' regions. Variants within these regions were flagged if they were predicted as pathogenic or likely pathogenic by *SnpEff* and *VEP* and went through multiple in silico prediction tools including *VarSeq* (Golen Helix™), *Ensembl* variant table, *gnomAD*, and *ExAC* report on selected variants, *Varsome*, and *VarAFT*. Also, *SWISS-MODEL* [20] and PyMol 2.4 [21] were applied to selected variants (predicted as highly damaging) for the implementation of homology modeling for confirmation of the negative effects of amino acid changes in the related protein.

PGx analysis with multiple dedicated bioinformatics tools

For analysis of variants in known and well-established PGx genes we used four PGx-dedicated tools: *Stargazer* (V.1.0.8), *Aldy* (V.3.3), *PharmCAT* (V.0.8.0), and *PharmaKU* which uses *Stargazer* V1.2.2 [22–26]. For *Aldy*, *PharmCAT*, and *PharmaKU*, GRCh38 BAM and VCF files were used and the tools were run according to their instructions in the accompanying documentation. *Stargazer* only works with GRCh37, hence the GRCh37-based VCF-only mode was used according to the documentation. All VCFs used in this section were the original files from the first standard analysis steps for NGS output containing all genes without any pre-filtering. Hence, results were obtained from all genes included in the tools which differed between tools. However, core pharmacogenes (defined as actionable in PGx guideline providers) were covered by all.

Haplotype/diplotype evaluation for well-known pharmacogenes

First, we made a comparison table for results from selected PGx-bioinformatics tools for nine core pharmacogenes: *CYP2B6, CYP2C19, CYP2C9, CYP2D6, CYP3A5, DPYD, SLCO1B1*, *UGT1A1*, and *VKORC1*. These genes are all present in guidelines from *PharmGKB*, *CPIC*, and *DPWG* in addition to being annotated in *PharmVar*. Results differed depending on the algorithm and variants used by each tool. A "3 vs. 1" conflict rule was used: if the same diplotype and phenotype were called by 3 out of the 4 tools, that was considered the correct assignment. If there was no majority agreement or if one tool did not give any calls, randomly selected discrepancies (to a total of 20 discrepancies) were manually investigated with the use of *PharmVar* to assess what the correct assignment was. The outcome resulted in the conclusion that *Stargazer* was most often correct. Hence, for discrepancies, the *Stargazer* assignment was selected as the correct call.

Possible drug-drug-gene interactions based on the final predicted metabolizer phenotypes or star alleles were theoretically assessed for all patients. For this, the registered demographic data and complete history of drugs plus clinical manifestations in the case of patients with reported ADRs' phenotype were used. Possible mismatches between the individuals' genotype-based prediction of drug metabolism and the true capacity to metabolize drugs were identified by freely available resources as a model of phenoconversion assessment for high throughput DNA sequencing data. Figure 1 illustrates our complete workflow for NGS-based clinical PGx tests for individuals.

Electronic health records and data storage

Haplotypes and phenotypes assigned in the previous step are included in the EHR in university hospital in Bialystok to guide future drug therapy for all participants. Additionally, results are reported back to the participants using a special PGx card as well. Such reporting methods are to allow the utilization of the information based on the provided guidelines or recommendations by *CPIC*, *FDA*, *DPWG*, or other guidelines. Each participant's profile in the current study includes specific records with information related to *CPIC* and *DPWG* guidelines plus novel variant data in less-known drug-related genes. Access to this database is provided through a publicly available Internet webpage, entitled "clinicalpgx.pl." (Figs. 2 and 3).

RESULTS

WES data analysis for clinical PGx practice

WES resulted in 1026.75 Gb of 101 bp paired-end reads the output and 93.92% of reads with $Q > 30$. On average, 30,000 variants were identified for each sample. PGx-VCF files displayed between 3300–3600 identified variants in 1800 drug-related genes for each sample. In total 299,297 unique variants from all samples passed the genotype quality and desired read depth (DP) filtrations. Out of the 299,297 variants, 4539 (1.51%) were identified as rare variants with minor allele frequency ≤ 0.01 based on data from the *1K genomes*, *gnomAD*, and *ExAC* databases. Also, the approach revealed 36 variants within our nine core pharmacogenes, with 28 of them considered rare and/or extremely rare in *1K genomes* and *gnomAD*. These 36 variants were not in *CPIC* or *PharmGKB*. Overall, of the 4539 rare variants, there were 21 frameshift, 19 in-frame deletion/insertion, 50 intronic, 18 splice-site, 26 stop codon, 1804 synonymous, and 2447 missense non-synonymous variants in coding regions plus 154 other types of changes (i.e., 27 UTR, 9 initial codon variants, etc.) found (Fig. 4). Multiple functional assessment algorithms identified 115 of the non-synonymous rare variants as damaging. The final step integrated with in silico analysis methods like extra deep filtration, deep computational analysis, and machine learning approaches alongside protein modeling implementation to inferring and providing higher accuracy rate in functional predictions for variants, particularly in less-known drug-related genes. Next, we checked the ability of common genetic bioinformatics tools (*SIFT*, *Polyphen2*, *FATHMM*, *Mutation taster*, *Mutation Assessor, and CAAD*) to identify known impactful variants in pharmacogenes. The evaluation analyzed a list of selected 39 interpreted variants (based on the U-PGx [27] consortiums panel which were previously analyzed through our other investigations and confirmed in PharmVAR) from 11 core pharmacogenes (*CYP2B6*, *CYP2C19*, *CYP2C9*, *CYP2D6*, *CYP3A5*, *DPYD*, *F5L*, *SLCO1B1*, *TPMT*, *UGT1A1*, and *VKORC1*). The results showed that most of the bioinformatics tools used were not successful in identifying these variants as potentially deleterious or impactful. This was particularly evident in the variants associated with a decrease or increase in function. However, variants that are known to be completely deleterious (loss of function) in PGx were identified in half of the cases. This was considered while we used *ExAC-LOF* as one of the main filtration tools for the detection of rare variants in our study. The tool contains information on loss of function variants from *ExAC*, which was one of our main databases for highlighting rare variants as well.

As the common bioinformatics tools like *SIFT*, *Polyphen2*, *CAAD*, etc. were shown to be not suitable for the identification of impactful PGx variants, we may do the pre-filtration of WES data for obtaining only PGx-related genes VCF file and use that in common tools. In this case, all the identified malfunction variants (for example loss of function, which are mostly highlighted by

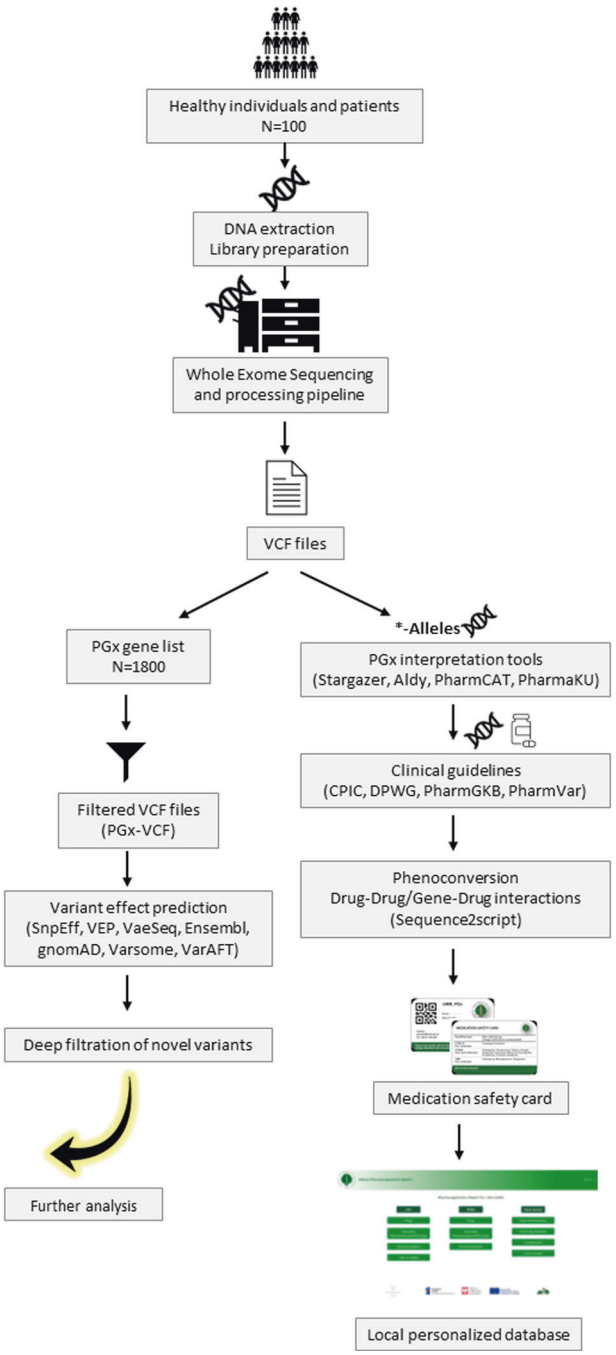

Fig. 1 Designed workflow for NGS-based comprehensive clinical PGx test data analysis. Obtained data from WES primary analysis divided into two main categories of variants from less-studied (not interpreted) and well-known pharmacogenes. For the less-studied gene, VCF files were filtered, using the Bed file for 1800 drug-related genes and undergone through deep computational functional assessment and variants effect prediction. Four PGx-dedicated bioinformatics tools are employed for Known genes star allele calling. Clinical guidelines were collected for identified markers in the previous step and samples' genotype and predicted phenotype were used in *sequence2script* platform for phenoconversion evaluations. Related PGx card created and actionable pharmacovariants for each participant stored in a secure database alongside recommendations from *CPIC* and *DPWG* plus information on novel variants. NGS next-generation sequencing, PGx pharmacogenomics, CPIC Clinical Pharmacogenetics Implementation Consortium, DPWG Dutch Pharmacogenomics Working Group.

common tools) come from drug-related genes making the downstream processing and detailed analysis more manageable. Then we can go further and do more in silico analysis by other available tools as the confirmation approach for predicted functional assessments.

Bioinformatics tools for NGS-PGx outcome (comparison of tools result)

Besides the prediction of yet unused variants in pharmacogenes, we have also explored the results from PGx-dedicated tools for the assignment of *-haplotypes based on well-known pharmacovariants. We evaluated these results in detail and evaluate the potential use of such tools for clinical practice. Four different software tools were used for assigning diplotypes (star alleles) and phenotypes. However, the result from these bioinformatics tools showed discrepancies. While *Aldy* and *Stargazer* showed the most similarity for nine core pharmacogenes, *PharmCAT* was unable to call star alleles for every gene. *PharmaKU* uses *Stargazer* as its basis (v V1.2.2) and accepts genome version GRCh38 as well [26]. While the outcome was mostly concordant with *Stargazer*, there were discrepancies (one per ten calls) mainly due to default calls in *PharmaKU*, especially in the absence of input data. If data on a gene or its variants was missing, *Stargazer* would not provide any results, *PharmaKU* on the other hand would assume the genes were entire wildtype and calls a *1/*1 haplotype. Figure 5 displays the concordance rate for PGx-dedicated bioinformatics tools and for each selected pharmacogene in detail. Also, Table 1 displays multi-tools' discrepancies reports and provides more details on calling star alleles. The overall result brought some important insights for these tool's functions, which are worth considering while using such tools in clinical PGx tests: (1) most common cause of discrepancies comes from differences in the variants each tool uses. Also, not all the tools implement phasing for haplotype detection. For example, while *Stargazer* uses Beagle as a built-in algorithm for running the phasing for the samples, *PharmCAT* works best with a phased VCF file as input data. *Aldy* uses only unphased data. (2) the genotype and phenotype assignments are not the same in every tool. For example, *Aldy's* variant to haplotype translation does not always match PharmVar (i.e., *CYP2B6**4 may call as *1). *PharmaKU*, on the other hand, does not provide any information on variants as a web-based report with only star alleles and predicted phenotype. Regarding phenotype assignments, *Stargazer's* phenotype predictions do not always match the guidelines (e.g., *CYP2B6**1/*6 is translated as a normal metabolizer instead of intermediate). Additionally, *Aldy* does not provide phenotype translation in the result, while the other evaluated tools have that. Finally, (3) the transparency and ease of use are different. Differences occur in all aspects of these tools: for example, the necessity for pre-processing of the input data (*Aldy*, *PharmCAT*, and complete mode for *Stargazer*), the comprehensiveness of the report, the technical features, and the genotype and phenotype translation.

Actionable pharmacovariants in individuals

The most "non-normal" phenotypes were identified for *CYP2D6. CYP3A5*, on the other hand, was the most consistent with almost all samples having a poor metabolizer phenotype. The overall frequency of abnormal alleles leads to aberrant phenotypes within our participants for nine core pharmacogenes were as follow: *CYP2B6* (47%), *CYP2C19* (17%), *CYP2C9* (31%), *CYP2D6* (60%), *CYP3A5* (90%), *DPYD* (6%), *SLCO1B1* (47%), *UGT1A1* (18%), and *VKORC1* (47%). Figure 6 indicates the frequency for each allele in selected genes and linked phenotypes in detail. Moreover, for running the theoretically phenoconversion measurements, particularly drug-drug-gene interactions on a large number of samples, we used *Sequence2script* [28] to identify any potential drug-drug-gene interactions. The assessment, however, showed almost no changes in drug response phenotypes and dosage modifications

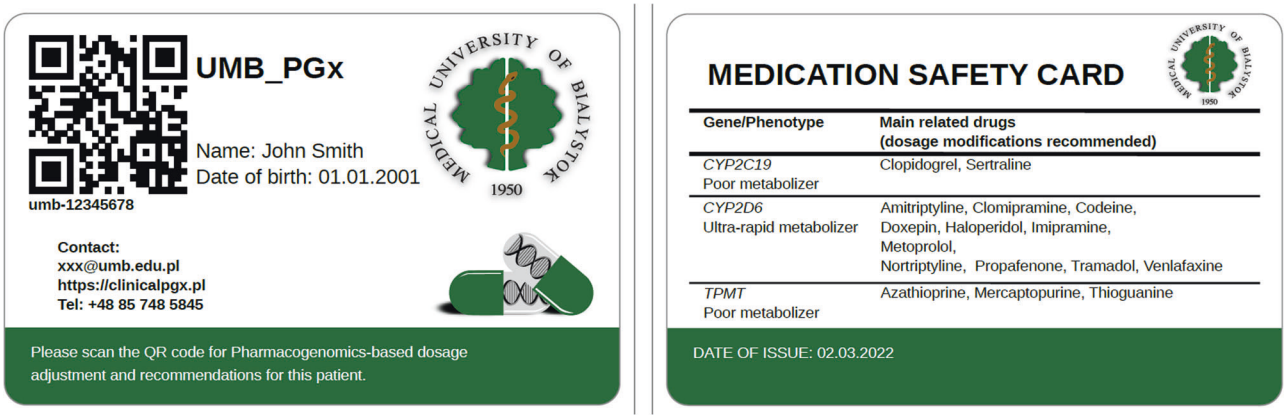

Fig. 2 Designed medication safety card for reaching out to individuals' clinical PGx test result. The card includes both unique number (for the physicians and pharmacists who do not access to QR reader) and QR code which are linked to the secure database (https://clinicalpgx.pl) for each person's PGx data. Core pharmacogenes with actionable variants for card holder listed in front of the card along with the main substrate which is needed to be considered while prescription for the person. PGx pharmacogenomics.

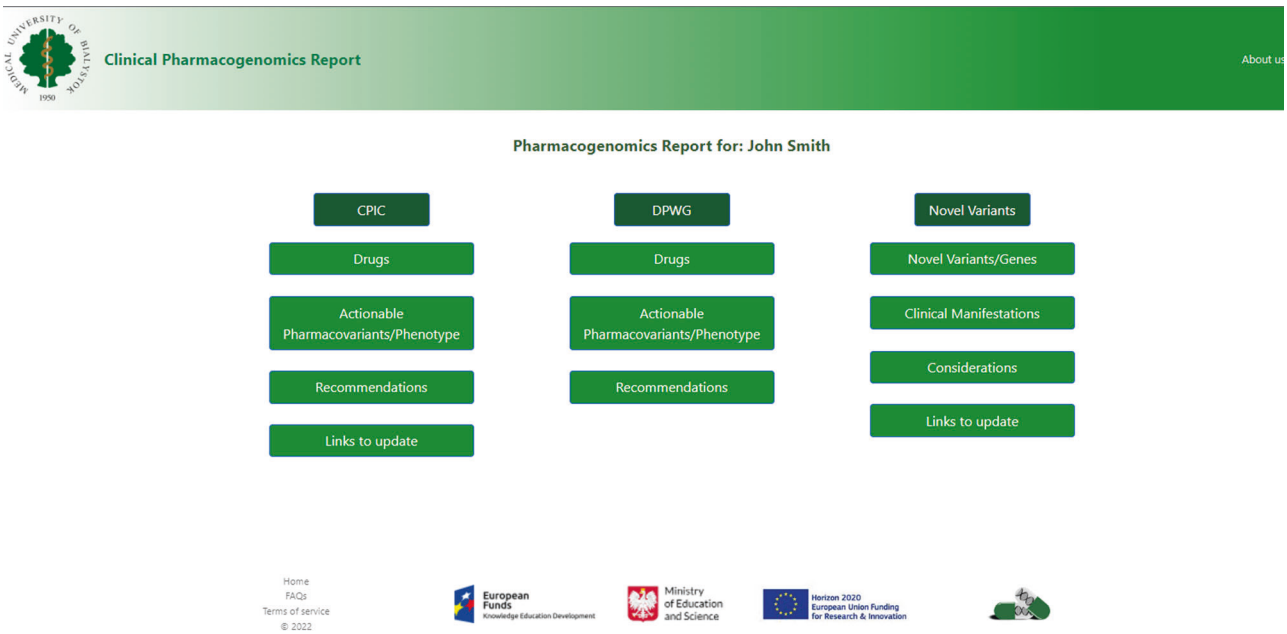

Fig. 3 Specific secure personalized database for individuals' clinical PGx test results and the recommendations from both *CPIC* and *DPWG* plus information on novel identified variants in drug-related genes. See the text for more details. PGx pharmacogenomics, CPIC Clinical Pharmacogenetics Implementation Consortium, DPWG Dutch Pharmacogenomics Working Group.

for our samples. Finally, all phenotypes are included on the PGx result card and on the website (clinicalpgx.pl), which contains both *CPIC* and *DPWG* guidelines to allow them to be accessed by the participants and their healthcare providers. An example of the result from our approach for actionable pharmacovariants and prescription recommendations is provided on "clinicalpgx.pl/data" for anonymous person.

DISCUSSION

The result of our study on PGx profiling from WES data may be helpful for utilization of bioinformatics tools, specifically PGx-dedicated algorithms, in daily clinical practice. Our investigation also displayed the advantage of pre-filtration of VCF files for only drug-related genes in order to help for identification of more pharmacovariants within such genes. We demonstrated higher accuracy of PGx independent bioinformatics tools, particularly in clinical research, compared to web-hosted algorithms like *PharmaKU* as the tool could be used just as a confirmation for *Stargazer* result, where the input data provided correctly. While the web-based PGx haplotype tools are easier to use, they also seem to be less accurate than the more transparent command line-based programs. In order to touch on the advantages of more in-depth methods in clinical reports, the selection and utilization of correct PGx-dedicated bioinformatics tool(s) must be considered by test centers as well. True applications of PGx bioinformatics algorithms in clinics will bring several advantages, not only in biomarker identification but also in physicians' accurate decision-making and drug stratification [29]. Choosing the right tool and annotation databases in addition to setting up a consistent workflow for routine practice in clinical centers requires advanced knowledge and awareness of existing tools or resources and their functional approaches in variant interpretations [30].

Even though all the available PGx-dedicated bioinformatics tools are limited to the specific number of pharmacogenes and included the distinct number of pharmacovariants in their panel,

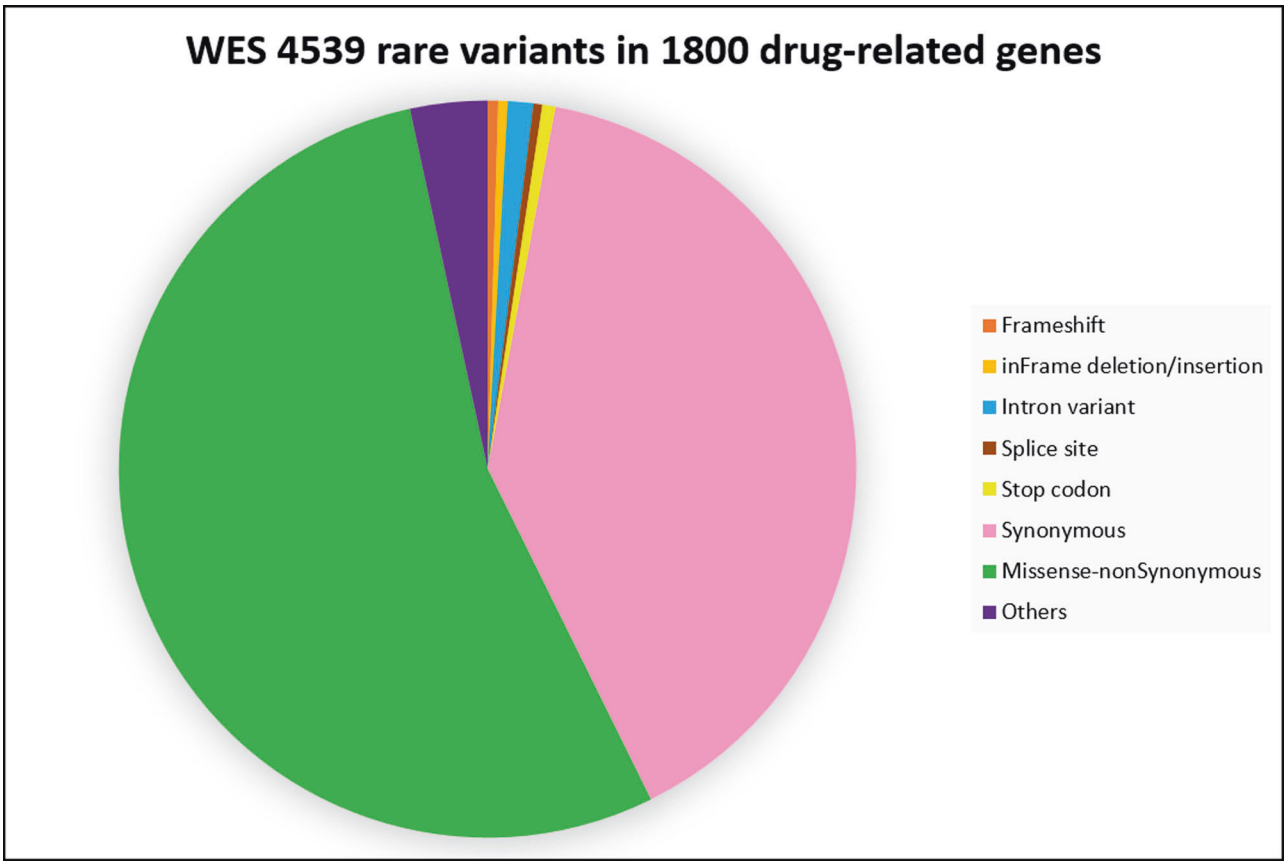

Fig. 4 Total rare variants in WES result for all samples. The distribution and functional impact of all rare variants in PGx-VCF files, which contain only drug-related genes. These variants went for the deep computational analysis, mostly for less-studied (not interpreted) drug-related genes. WES whole-exome sequencing, PGx pharmacogenomics.

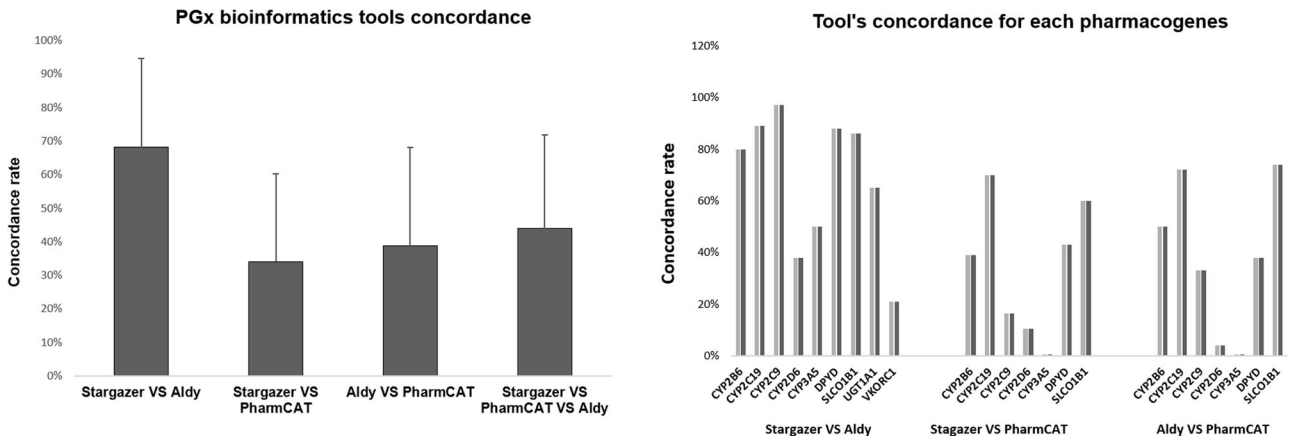

Fig. 5 PGx-dedicated bioinformatics algorithms concordance rate for selected nine core pharmacogenes in the current study. The result for the total comparison of tools is illustrated as well. The concordance rate was calculated when there was at least one call for the variants. *PharmaKU* is not included here as it uses *Stargazer* as a built-in algorithm and the result was mainly the same when the correct input data was used (see the main text for more details). *PharmCAT* did not call any alleles for *UGT1A1* and *VKORC1*. PGx pharmacogenomics.

most of these curated genetic variations come with clinical guidelines and annotations for treatment modifications as well. Hence, applying multi-tools for including more genes and variants seems reasonable. So far, most studies reported the advantages of using PGx bioinformatics tools in clinical investigations but as a separate entity [31, 32]. For those reported the multi-tool utilization, again not all of the genes in all samples were evaluated in that way [33]. Among PGx-dedicated bioinformatics algorithms, we propose to use at least two of such tools for providing more confident haplotype calls. However, an important limitation of applying different tools would be the necessity for running the alignment part for different reference genomes as some of them might need GRCh37 while the others work with GRCh38. In our study, we tried to use NCBI's genome remapping service (https://www.ncbi.nlm.nih.gov/genome/tools/remap) to perform this re-alignment and liftover from GRCh38 to GRCh37. However, after evaluating of results, we noticed that the approach led to the exclusion of some important PGx variants. Therefore, we decided

Table. 1. Multi PGx-dedicated bioinformatics tools' discrepancies report for each core pharmacogene in current study and the total measurement for such conflicts in 100 WES data.

Genes	*CYP2B6*	*CYP2C19*	*CYP2C9*	*CYP2D6*	*CYP3A5*	*DPYD*	*SLCO1B1*	*UGT1A1*	*VKORC1*
Same result in all tools	70	76	98	8	23	91	43	63	100
3 vs. 1*	15	11	0	1	1	0	45	0	0
2 vs. 2	4	2	1	0	0	0	7	0	0
2 vs. 1 vs. 1	5	2	0	12	0	0	4	0	0
2 vs. 1 (One tool did not call any diplotype)	6	9	1	40	75	6	0	19	Not applicable
All different	0	0	0	39	1	3	1	18	Not applicable
Total conflicts without 3 vs. 1	15	13	2	91	76[b]	9	12	37	–

*Only "3 vs. 1" was not checked for further evaluations. No matched phenotype was removed from the final report. Tools' report files for the rest of the "vs." situations are checked manually against *PharmVAR* and *PharmGKB*. Wrong or non-clear calls are interpreted as not accepted calls and removed. The overall concordance rate for all tools: 71% (including 3 vs. 1 scenario) see the main text for more details.
[a]*Stargazer* VCF only mode for calling stars in *CYP2D6* as a highly structural polymorphic gene is not preferred.
[b]*CYP3A5* alleles are defined in a different way in *Stargazer* and *PharmaKU*.
[c]Tools in *SLCO1B1* (less), *UGT1A1*, and *VKORC1* use different allele nomenclature. Hence, the major discrepancies came from different allele names.

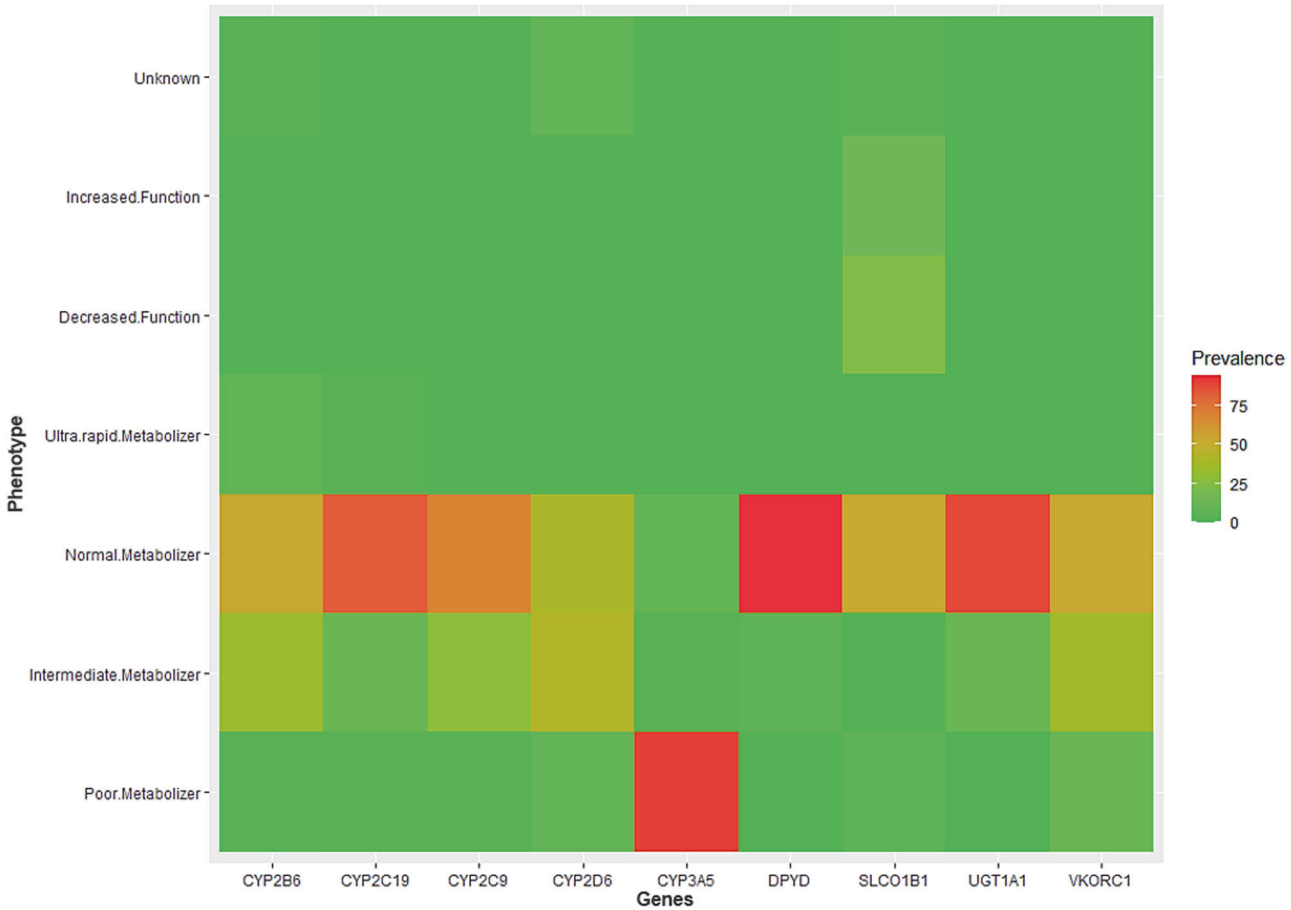

Fig. 6 Different metabolizer for nine core pharmacogenes. Distribution and prevalence of different metabolizer phenotypes and related alleles for selected genes in the current study.

to perform a separate realignment with GRCh37 assembly as the reference for the raw data. These data were subsequently used by *Stargazer*.

The outcome of our result in the adaptation of PGx-dedicated bioinformatics tools for clinical interpretation of PGx variants may help clinicians to improve the implementation of the NGS-guided clinical PGx tests. Once the utilization of such computational assessments is established in the center, the related healthcare system may benefit from the fast and more accurate PGx marker diagnosis in a shorter turnaround time.

Also, it is worth considering that not all types of PGx variants may be identified by common bioinformatics tools. As we have shown, increased and decreased functions (rapid and intermediate metabolizers) are mostly ignored by tools like *SIFT*, *Polyphen2*, *FATHMM*, *Mutation taster*, *Mutation Assessor*, etc. Therefore, it might be valuable to filter PGx regions from VCF files to select

candidates for in silico validation studies as opposed to using inaccurate in silico tools for the assessment of variant impact. This type of approach would be for novel variants with unknown significance, which have an impact on the protein (e.g., missense or frameshift variants) [34]. Nevertheless, the workflow in this level for the current study brought many interesting outcomes for novel and/or not-annotated variants in our samples, which the interpretation and further analysis are still in progress. Today, computational assessments proved to be a promising approach for the translation of novel variants into healthcare [35, 36].

Besides gene-drug interactions it is also of importance to consider phenoconversion for improving accuracy rate for phenotype prediction in personalized therapy area too. Under the influence of comedication the activity of an enzyme can switch, for example from intermediate to normal metabolizer due to an inducer effect. For instance, proton-pump inhibitors can reduce CYP2C19 activity and thereby convert a normal metabolizer phenotype to an intermediate metabolizer phenotype. This can, in turn, has an impact on other drugs used by the patients. The use of both proton-pump inhibitors (CYP2C19 inhibitor) and clopidogrel (CYP2C19 substrate) is highly likely in a cardiovascular cohort such as ours, therefore it is important to be aware of these types of drug-drug-gene interactions. Future research and investigations will need to take comedication into account when studying the impact of PGx on clinical outcomes.

Integration of computational assessment and bioinformatic functional analysis of pharmacovariants within high throughput DNA sequencing data is rapidly expanding. However, while a comprehensive clinical PGx profile could be successfully extracted from WES data, available tools to interpret such data are not consistent for all pharmacogenes and show several discrepancies compared to each other. Moreover, WES data demonstrates an abundance of variants not yet used in clinical practice. To bring the translation of such technologies into daily clinical setting the clinical validity and utility of dedicated bioinformatics tools should be investigated more.

DATA AVAILABILITY

The data are available from the corresponding author at Department of Analysis and Bioanalysis of Medicines, Medical University of Bialystok on reasonable request.

REFERENCES

1. Abdullah-Koolmees H, Van Keulen AM, Nijenhuis M, Deneer VH. Pharmacogenetics guidelines: overview and comparison of the DPWG, CPIC, CPNDS, and RNPGx guidelines. Front Pharmacol. 2021;11:1976.
2. Scott SA. Clinical pharmacogenomics: opportunities and challenges at point of care. Clin Pharmacol Ther. 2013;93:33–5.
3. Kozyra M, Ingelman-Sundberg M, Lauschke VM. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. Genet Med. 2017;19:20–9.
4. Lopes JL, Harris K, Karow MB, Peterson SE, Kluge ML, Kotzer KE, et al. Targeted genotyping in clinical pharmacogenomics: what is missing? J Mol Diagn. 2022;24:253–61.
5. Yang Y, Botton MR, Scott ER, Scott SA. Sequencing the CYP2D6 gene: from variant allele discovery to clinical pharmacogenetic testing. Pharmacogenomics. 2017;18:673–85.
6. van der Lee M, Allard WG, Bollen S, Santen GW, Ruivenkamp CA, Hoffer MJ, et al. Repurposing of diagnostic whole exome sequencing data of 1,583 individuals for clinical pharmacogenetics. Clin Pharmacol Ther. 2020;107:617–27.
7. Cousin MA, Matey ET, Blackburn PR, Boczek NJ, McAllister TM, Kruisselbrink TM, et al. Pharmacogenomic findings from clinical whole exome sequencing of diagnostic odyssey patients. Mol Genet Genom Med. 2017;5:269–79.
8. Santos M, Niemi M, Hiratsuka M, Kumondai M, Ingelman-Sundberg M, Lauschke VM, et al. Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. Genet Med. 2018;20:622–9.
9. Price MJ, Carson AR, Murray SS, Phillips T, Janel L, Tisch R, et al. First pharmacogenomic analysis using whole exome sequencing to identify novel genetic determinants of clopidogrel response variability: results of the genotype information and functional testing (GIFT) exome study. J Am Coll Cardiol. 2012;59:E9–E.
10. Mizzi C, Peters B, Mitropoulou C, Mitropoulos K, Katsila T, Agarwal MR, et al. Personalized pharmacogenomics profiling using whole-genome sequencing. Pharmacogenomics. 2014;15:1223–34.
11. Sivadas A, Salleh M, Teh L, Scaria V. Genetic epidemiology of pharmacogenetic variants in South East Asian Malays using whole-genome sequences. Pharmacogenomics J. 2017;17:461–70.
12. Choi J, Tantisira KG, Duan QL. Whole genome sequencing identifies high-impact variants in well-known pharmacogenomic genes. Pharmacogenomics J. 2019;19:127–35.
13. Ng D, Hong CS, Singh LN, Johnston JJ, Mullikin JC, Biesecker LG. Assessing the capability of massively parallel sequencing for opportunistic pharmacogenetic screening. Genet Med. 2017;19:357–61.
14. Gordon A, Fulton RS, Qin X, Mardis ER, Nickerson DA, Scherer S. PGRNseq: a targeted capture sequencing panel for pharmacogenetic research and implementation. Pharmacogenet Genomics. 2016;26:161.
15. Gulilat M, Lamb T, Teft WA, Wang J, Dron JS, Robinson JF, et al. Targeted next generation sequencing as a tool for precision medicine. BMC Med Genomics. 2019;12:1–17.
16. Tafazoli A, Guchelaar H-J, Miltyk W, Kretowski AJ, Swen JJ. Applying next-generation sequencing platforms for pharmacogenomic testing in clinical practice. Front Pharmacol. 2021;12:2025.
17. Twesigomwe D, Wright GE, Drögemöller BI, da Rocha J, Lombard Z, Hazelhurst S. A systematic comparison of pharmacogene star allele calling bioinformatics algorithms: a focus on CYP2D6 genotyping. NPJ Genom Med. 2020;5:1–11.
18. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6:80–92.
19. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17:1–14.
20. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46:W296–303.
21. DeLano WL. The PyMOL molecular graphics system. 2002. http://www.pymol.org.
22. Lee S-B, Wheeler MM, Patterson K, McGee S, Dalton R, Woodahl EL, et al. Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. Genet Med. 2019;21:361–72.
23. Lee SB, Wheeler MM, Thummel KE, Nickerson DA. Calling star alleles with stargazer in 28 pharmacogenes with whole genome sequences. Clin Pharmacol Ther. 2019;106:1328–37.
24. Numanagić I, Malikić S, Ford M, Qin X, Toji L, Radovich M, et al. Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. Nat Commun. 2018;9:1–11.
25. Sangkuhl K, Whirl-Carrillo M, Whaley RM, Woon M, Lavertu A, Altman RB, et al. Pharmacogenomics clinical annotation tool (Pharm CAT). Clin Pharmacol Ther. 2020;107:203–10.
26. John SE, Channanath AM, Hebbar P, Nizam R, Thanaraj TA, Al-Mulla F. PharmaKU: a web-based tool aimed at improving outreach and clinical utility of pharmacogenomics. J Pers Med. 2021;11:210.
27. van der Wouden CH, van Rhenen MH, Jama WO, Ingelman-Sundberg M, Lauschke VM, Konta L, et al. Development of the PG x-Passport: a panel of actionable germline genetic variants for pre-emptive pharmacogenetic testing. Clin Pharmacol Ther. 2019;106:866–73.
28. Bousman CA, Wu P, Aitchison KJ, Cheng T. Sequence2Script: a web-based tool for translation of pharmacogenetic data into evidence-based prescribing recommendations. Front Pharmacol. 2021;12:238.
29. Overkleeft R, Tommel J, Evers AW, den Dunnen JT, Roos M, Hoefmans M-J, et al. Using personal genomic data within primary care: a bioinformatics approach to pharmacogenomics. Genes. 2020;11:1443.
30. Tong H, Phan NV, Nguyen TT, Nguyen DV, Vo NS, Le L. Review on databases and bioinformatic approaches on pharmacogenomics of adverse drug reactions. Pharmacogenomics Pers Med. 2021;14:61.
31. Ly RC, Shugg T, Ratcliff R, Osei W, Lynnes TC, Pratt VM, et al. Analytical validation of a computational method for pharmacogenetic genotyping from clinical whole exome sequencing. J Mol Diagn. 2022;24:576–85.
32. Wankaew N, Chariyavilaskul P, Chamnanphon M, Assawapitaksakul A, Chetruengchai W, Pongpanich M, et al. Genotypic and phenotypic landscapes of 51 pharmacogenes derived from whole-genome sequencing in a Thai population. PLoS ONE. 2022;17:e0263621.

33. Caspar SM, Schneider T, Meienberg J, Matyas G. Added value of clinical sequencing: WGS-based profiling of pharmacogenes. Int J Mol Sci. 2020;21:2308.
34. Siamoglou S, Koromina M, Hishinuma E, Yamazaki S, Tsermpini E-E, Kordou Z, et al. Identification and functional validation of novel pharmacogenomic variants using a next-generation sequencing-based approach for clinical pharmacogenomics. Pharmacol Res. 2022;176:106087.
35. Zhou Y, Mkrtchian S, Kumondai M, Hiratsuka M, Lauschke VM. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. Pharmacogenomics J. 2019;19:115–26.
36. Pandi M-T, Koromina M, Tsafaridis I, Patsilinakos S, Christoforou E, van der Spek PJ, et al. A novel machine learning-based approach for the computational functional assessment of pharmacogenomic variants. Hum Genomics. 2021;15:1–13.

AUTHOR CONTRIBUTIONS

AT designed the study, performed most of the lab work and ran both common and PGx-dedicated bioinformatics algorithms, did WES data interpretations, analyzed the PGx-related tools data, designed the webpage for clinical PGx data and medication safety card, and wrote the entire manuscript draft, MvdL helped in running PGx-dedicated bioinformatics algorithms and analyzing the related data, and edit the manuscript, AZ and NWK helped and ran some parts of lab work, HM and RHPV did the raw WES data analysis for 100 samples, KK and AZ created and launched the webpage for clinical PGx data plus medication safety card, JJS and WM supervised the entire study and edited final version of the manuscript.

FUNDING

This article was conducted within the projects which have received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant, agreement no. 754432, and the Polish Ministry of Science and Higher Education and from financial resources for science in 2018–2023 granted for the implementation of an international co-financed project.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41397-022-00286-4.

Correspondence and requests for materials should be addressed to Wojciech Miltyk.

Reprints and permission information is available at http://www.nature.com/reprints

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

10. Summary

– Summary in Polish:

Wyniki badań przedstawione w niniejszej rozprawie doktorskiej dotyczą zastosowania oprogramowania bioinformatycznego związanego z farmakogenomiką (PGx) oraz wykorzystania nowych sposobów analizy mniej znanych farmakogenów i farmakowariantów. Cztery publicznie dostępne algorytmy bioinformatyczne PGx do analizy haplotypów PGx zostały zastosowane do oceny dziewięciu wybranych, bardzo ważnych farmakogenów (VIP) wykazując wysoki (45-70%) współczynnik zgodności. Ponadto w celu przeprowadzenia oceny funkcjonalnej niezinterpretowanych farmakowariantów zastosowano metody głębokiej filtracji obliczeniowej wielkoskalowego profilowania klinicznego PGx. Aby zapewnić lekarzom i farmaceutom dostęp do uzyskanych wyników famakogenomicznych u badanych pacjentów, zidentyfikowane warianty zostały umieszczone w opracowanej bazie danych na stronie internetowej (pierwszej i dotychczas jedynej w Polsce), oraz opracowano pacjentom zindywidualizowane karty PGx zawierające dane o ich zidentyfikowanych wariantach farmakogenów (pierwsze w Polsce), co pozwala na szybki dostęp do wyników badań i indywidualizację ich trerapii.

– Summary in English:

Main result of studies included in this dissertation are about PGx-related bioinformatic software outcomes and the approaches to deal with less-studied pharmacogenes and pharmacovariants. Four publicly available PGx bioinformatics algorithms to assign PGx haplotypes were applied to nine selected very important pharmacogenes (VIP) and revealed a 45–70% concordance rate. To ensure availability of the results at point-of-care, actionable variants were stored in a web-hosted database and PGx-cards were developed for quick access and handed to the study subjects. Also, methods for deep computational filtration of large scale clinical PGx profiling introduced in order to perform functional assessment of not-interpreted pharmacovariants.

11. Bibliography

1. Reisberg S, Krebs K, Lepamets M, Kals M, Mägi R, Metsalu K, Lauschke VM, Vilo J, Milani L. Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. Genetics in Medicine. 2019 Jun;21(6):1345-54.
2. Ingelman-Sundberg M, Mkrtchian S, Zhou Y, Lauschke VM. Integrating rare genetic variants into pharmacogenetic drug response predictions. Human genomics. 2018 Dec;12(1):1-2.
3. Wake DT, Smith DM, Kazi S, Dunnenberger HM. Pharmacogenomic clinical decision support: a review, how-to guide, and future vision. Clinical Pharmacology & Therapeutics. 2022 Jul;112(1):44-57.
4. Zhou Y. Assessing the Importance of Rare Genetic Variants for Drug Response. Karolinska Institutet (Sweden); 2021.
5. Krebs K, Milani L. Translating pharmacogenomics into clinical decisions: do not let the perfect be the enemy of the good. Human genomics. 2019 Dec;13(1):1-3.
6. van der Wouden CH, Bank PC, Özokcu K, Swen JJ, Guchelaar HJ. Pharmacist-initiated pre-emptive pharmacogenetic panel testing with clinical decision support in primary care: record of PGx results and real-world impact. Genes. 2019 May 29;10(6):416.
7. Huser V, Cimino JJ. Providing pharmacogenomics clinical decision support using whole genome sequencing data as input. AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science. 2013 Mar 18;2013:81-.
8. Abdullah-Koolmees H, Van Keulen AM, Nijenhuis M, Deneer VH. Pharmacogenetics guidelines: overview and comparison of the DPWG, CPIC, CPNDS, and RNPGx guidelines. Frontiers in pharmacology. 2021 Jan 25;11:595219.
9. Zhou Y, Mkrtchian S, Kumondai M, Hiratsuka M, Lauschke VM. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. The pharmacogenomics journal. 2019 Apr;19(2):115-26.
10. van der Lee M, Kriek M, Guchelaar HJ, Swen JJ. Technologies for pharmacogenomics: a review. Genes. 2020 Dec 4;11(12):1456.
11. Devarajan S, Moon I, Ho MF, Larson NB, Neavin DR, Moyer AM, Black JL, Bielinski SJ, Scherer SE, Wang L, Weinshilboum RM. Pharmacogenomic next-generation DNA sequencing: lessons from the identification and functional characterization of variants of unknown significance in CYP2C9 and CYP2C19. Drug Metabolism and Disposition. 2019 Apr 1;47(4):425-35.

12. Arbitrio M, Scionti F, Di Martino MT, Caracciolo D, Pensabene L, Tassone P, Tagliaferri P. Pharmacogenomics biomarker discovery and validation for translation in clinical practice. Clinical and Translational Science. 2021 Jan;14(1):113-9.
13. Pandi MT, Koromina M, Tsafaridis I, Patsilinakos S, Christoforou E, van der Spek PJ, Patrinos GP. A novel machine learning-based approach for the computational functional assessment of pharmacogenomic variants. Human genomics. 2021 Dec;15(1):1-3.
14. Lee M. Towards solving the missing heritability in pharmacogenomics. (2022).
15. Blagec K, Swen JJ, Koopmann R, Cheung KC, Crommentuijn-van Rhenen M, Holsappel I, Konta L, Ott S, Steinberger D, Xu H, Cecchin E. Pharmacogenomics decision support in the U-PGx project: Results and advice from clinical implementation across seven European countries. PloS one. 2022 Jun 8;17(6):e0268534.
16. Caspar SM, Schneider T, Meienberg J, Matyas G. Added value of clinical sequencing: WGS-based profiling of pharmacogenes. International journal of molecular sciences. 2020 Mar 26;21(7):2308.
17. Giannopoulou E, Katsila T, Mitropoulou C, Tsermpini EE, Patrinos GP. Integrating next-generation sequencing in the clinical pharmacogenomics workflow. Frontiers in pharmacology. 2019 Apr 5;10:384.
18. Zhou Y, Fujikura K, Mkrtchian S, Lauschke VM. Computational methods for the pharmacogenetic interpretation of next generation sequencing data. Frontiers in Pharmacology. 2018 Dec 4;9:1437.
19. Lee EM, Xu K, Mosbrook E, Links A, Guzman J, Adams DR, Flynn E, Valkanas E, Toro C, Tifft CJ, Boerkoel CF. Pharmacogenomic incidental findings in 308 families: The NIH Undiagnosed Diseases Program experience. Genetics in Medicine. 2016 Dec;18(12):1303-7.
20. van der Wouden CH, van Rhenen MH, Jama WO, Ingelman-Sundberg M, Lauschke VM, Konta L, Schwab M, Swen JJ, Guchelaar HJ. Development of the PG x-Passport: A Panel of Actionable Germline Genetic Variants for Pre-Emptive Pharmacogenetic Testing. Clinical Pharmacology & Therapeutics. 2019 Oct;106(4):866-73.
21. Overkleeft R, Tommel J, Evers AW, den Dunnen JT, Roos M, Hoefmans MJ, Schrader WE, Swen JJ, Numans ME, Houwink EJ. Using personal genomic data within primary care: a bioinformatics approach to pharmacogenomics. Genes. 2020 Nov 30;11(12):1443.

12. Authorship Contribution Statements

(Information about nature of participation and author's contribution (in %) and statements from co-authors)

1. Applying next-generation sequencing platforms for Pharmacogenomics testing in clinical practice, *Front. Pharmacol, Pharmacogenetics and Pharmacogenomics section*. (2021) DOI: 10.3389/fphar.2021. (12) - 693453

Author's name and surmane	Nature of participation	Contribution in %
Ph.D. Student: Alireza Tafazoli	Designed the study, conducted the search for literature, and wrote the entire manuscript.	65%
Prof. Henk Jan Guchelaar	Supervised the study thoroughly and revised and edited the manuscript.	10%
Prof. Adam Kretowski	Provided the idea and introduced the topic for the manuscript.	5%
Prof. Wojciech Miltyk	Performed the search for literature and modified the text as well.	10%
Dr. Jesse J. Swen	Supervised the study thoroughly and revised and edited the manuscript.	10%

2. Pharmacogenomics, how to deal with different types of variants in next generation sequencing data in the personalized medicine area, *J. Clin. Med.* (2021) DOI: 10.3390/jcm10010034.

Author's name and surmane	Nature of participation	Contribution in %
Ph.D. Student: Alireza Tafazoli	Provided the idea, designed the study, conducted the literature search, and wrote the entire manuscript.	74%
Dr. Natalia W.Kurylonek	Searched for the literature and edited the manuscript.	8%
Dr. Renata Posmyk	Searched for the literature and edited the manuscript.	8%
Prof. Wojciech Miltyk	Supervised the study thoroughly and revised the manuscript.	10%

3. Development of an extensive workflow for comprehensive clinical Pharmacogenomic profiling: *lessons from a pilot study on 100 Whole Exome Sequencing data.* Pharmacogenomics J. (2022) DOI: 10.1038/s41397-022-00286-4

Author's name and surmane	Nature of participation	Contribution in %
Ph.D. Student: Alireza Tafazoli	Designed the study, performed most of the lab work and ran PGx dedicated bioinformatics algorithms, did WES data interpretations, analysed the PGx-related tools data, designed the webpage for clinical PGx data and	50%

	medication safety card, and wrote the entire manuscript draft.	
Dr. Maaike van der Lee	Helped in running PGx dedicated bioinformatics algorithms and analysing the related data, and edited the manuscript.	10%
Dr. Jesse J. Swen	Supervised the entire study and edited final version of the manuscript.	10%
mgr. Anna Zeller	Helped and ran some parts of lab work.	3%
Dr. Natalia W.Kurylonek	Helped and ran some parts of lab work.	3%
Dr. Hailiang Mei	Did the raw WES data analysis for 100 samples.	3%
Dr. Ruben H.P. Vorderman	Did the raw WES data analysis for 100 samples.	3%
Dr. Krzysztof Konopko	Created and launched the webpage for clinical PGx data plus medication safety card.	3%
Dr. Andrzej Zankiewicz	Created and launched the webpage for clinical PGx data plus medication safety card.	5%
Prof. Wojciech Miltyk	Supervised the entire study and edited final version of the manuscript.	10%

I hereby declare that all co-authors agreed to use these articles in the current dissertation.

Signature

Ph.D. project related published papers:

1. **Tafazoli, A.,** van der Lee, M., Swen, J.J., Zeller, A., Wawrusiewicz-Kurylonek, N., Mei, H., Vorderman, R., Konopko, K., Zankiewicz, A., and Miltyk, W., (2022). Development of an extensive workflow for comprehensive clinical Pharmacogenomic profiling: lessons from a pilot study on 100 Whole Exome Sequencing data. *Pharmacogenomics J.* DOI: 10.1038/s41397-022-00286-4
2. **Tafazoli, A.**, Guchelaar, H.J., Kretowski, A., Miltyk, W, Swen, J.J. (2021). Applying next-generation sequencing platforms for Pharmacogenomics testing in clinical practice, *Front. Pharmacol, Pharmacogenetics and Pharmacogenomics section*. DOI: 10.3389/fphar.2021.693453
3. **Tafazoli, A.**, Guggilla, R.K., Kamel-Koleti, Z., Miltyk, W. (2021). Strategies to improve the clinical outcomes for direct-to-consumer Pharmacogenomics tests (DTC-PT), *Genes*. Doi: 10.3390/genes12030361
4. **Tafazoli, A.**, Wołczyński, S., Kurylonek, N.W., Esmaeili, S.A., Miltyk, W. (2021). Pharmacogenomics Biomarkers of Follicle-Stimulating Hormone Receptor Malfunction in Females with Impaired Ovarian Response- *a genetic survey*, *J. Clin. Med*. Doi: 10.3390/jcm10020170
5. **Tafazoli, A.**, Kurylonek, N.W., Posmyk, R., Miltyk, W. (2021). Pharmacogenomics, how to deal with different types of variants in next generation sequencing data in the personalized medicine area, *J. Clin. Med*. Doi: 10.3390/jcm10010034
6. Makvandi, P., Ghomi, M., Ashrafizadeh, M., **Tafazoli, A.**, Agarwal, T., Delfi, M., Akhtari, J., Nazarzade Zare, E., Padil, V.T., Zarrabi. A., Pourreza, N., Miltyk, W., Maiti, T.K., (2020). A review on advances in graphene-derivative/polysaccharide bionanocomposites: Therapeutics, Pharmacogenomics and toxicity. *Carbohydr. Polym*. 22:116952. Doi: 10.1016/j.carbpol.2020.116952.

Alireza Tafazoli, all publications:

International with peer review
8 as first author, 3 as 2nd author, 9 as co-author

1) **Tafazoli, A.**, van der Lee, M., Swen, J., Zeller, A., Wawrusiewicz-Kurylonek, N., Mei, H., Vorderman, R., Konopko, K., Zankiewicz, A., and Miltyk, W., (2022). Development of an extensive workflow for comprehensive clinical Pharmacogenomic profiling: *lessons from a pilot study on 100 Whole Exome Sequencing data*. *Pharmacogenomics J.* DOI: 10.1038/s41397-022-00286-4
2) **Tafazoli, A.**, Guchelaar, H.J., Kretowski, A., Miltyk, W, Swen, J.J. (2021). Applying next-generation sequencing platforms for Pharmacogenomics testing in clinical practice, *Front. Pharmacol, Pharmacogenetics and Pharmacogenomics section*. DOI: 10.3389/fphar.2021.693453
3) **Tafazoli, A.**, Guggilla, R.K., Kamel-Koleti, Z., Miltyk, W. (2021). Strategies to improve the clinical outcomes for direct-to-consumer Pharmacogenomics tests (DTC-PT), *Genes*. Doi: 10.3390/genes12030361
4) **Tafazoli, A.**, Wołczyński, S., Kurylonek, N.W., Esmaeili, S.A., Miltyk, W. (2021). Pharmacogenomics Biomarkers of Follicle-Stimulating Hormone Receptor Malfunction in Females with Impaired Ovarian Response- *a genetic survey*, *J. Clin. Med*. Doi: 10.3390/jcm10020170
5) **Tafazoli, A.**, Kurylonek, N.W., Posmyk, R., Miltyk, W. (2021). Pharmacogenomics, how to deal with different types of variants in next generation sequencing data in the personalized medicine area, *J. Clin. Med*. Doi: 10.3390/jcm10010034
6) Khoshhal, H., Tabari, M.A., **Tafazoli, A.**, Khandan, M., Bagheri, A. (2020). Applying Computer Simulations Data in Battling with COVID-19, *using pre-analyzed molecular and chemical data to face pandemic*, *Informatics in Medicine Unlocked (IMU)*. 17:100458. Doi: 10.1016/j.imu.2020.100458.
7) Parizi, P.K., Yarahmadi, F., Movahedi Tabar, H., Hosseini, Z., Sarli, A., **Tafazoli, A.**, Esmaeili, A., (2020). miRNAs and Target Molecules in Bladder Cancer, *Medical Oncology*. 37(12):1-33. Doi: 10.1007/s12032-020-01435-0.
8) Makvandi, P., Ghomi, M., Ashrafizadeh, M., **Tafazoli, A.**, Agarwal, T., Delfi, M., Akhtari, J., Nazarzade Zare, E., Padil, V.T., Zarrabi. A., Pourreza, N., Miltyk, W., Maiti, T.K., (2020). A review on advances in graphene-derivative/polysaccharide bionanocomposites: Therapeutics, Pharmacogenomics and toxicity. *Carbohydr. Polym*. 22:116952. Doi: 10.1016/j.carbpol.2020.116952.
9) **Tafazoli, A**., Behjati, F., Farhud, DD., Abbaszadegan, MR. (2019). Combination of Genetics and Nanotechnology for Down syndrome Modification: A Potential Hypothesis and Review of the Literature. *IJPH*, Mar 12;48(3):371-8.
10) Akbari Kordkheyli V, Khonakdar Tarsi A, Mishan MA, **Tafazoli A**, Bardania H, Zarpou S, Bagheri A. (2019). Effects of quercetin on microRNAs: A mechanistic review. *J. Cell. Biochem*. 120(8):12141-12155, Doi: 10.1002/jcb.28663
11) Zeynalzadeh, M., **Tafazoli, A**., Aarabi, A., Moghaddassian, M., Ashrafzadeh, F., Houshmand, M., Taghehchian, N., & Abbaszadegan, M. R. (2018). Four novel mutations of the *BCKDHA*, *BCKDHB* and *DBT* genes in Iranian patients with maple syrup urine disease. *J Pediatr Endocrinol Metab*, *31*(2), 205-212. Doi:10.1515/jpem-2017-0305
12) **Tafazoli, A**., Eshraghi, P., Pantaleoni, F., Vakili, R., Moghaddassian, M., Ghahraman, M., Muto, V., Paolacci, S., Golyan, F. F., & Abbaszadegan, M. R. (2018). Novel mutations and

their genotype-phenotype correlations in patients with Noonan syndrome, using next-generation sequencing. *Adv Med Sci, 63*(1), 87-93. Doi:10.1016/j.advms.2017.07.001

13) Ebrahimzadeh, M. A., **Tafazoli, A**., Akhtari, J., Biparva, P., & Eslami, S. (2018). Engineered silver nanoparticles, a new nanoweapon against cancer. *Anticancer Agents Med Chem*. Doi:10.2174/1871520618666180808093040
14) Akhtari, J., **Tafazoli, A**., Mehrad-Majd, H., & Mahrooz, A. (2018). Nanovehicle-based small interfering RNA (siRNA) delivery for therapeutic purposes: a new molecular approach in the Pharmacogenomics. *Curr Clin Pharmacol*. Doi:10.2174/1574884713666180709152610
15) **Tafazoli, A**., Eshraghi, P., Kamel-Koleti, Z., & Abbaszadegan, M. (2017). Noonan syndrome - a new survey. *Arch Med Sci, 13*(1), 215-222. Doi:10.5114/aoms.2017.64720
16) Hashemi-Soteh, M. B., Mousavi, S. S., & **Tafazoli, A**. (2017). Haplotypes inside the beta-globin gene: use as new biomarkers for beta-thalassemia prenatal diagnosis in north of Iran. *J Biomed Sci, 24*(1), 92. Doi:10.1186/s12929-017-0396-y
17) Khaledi, A., Bahador, A., Esmaeili, D., **Tafazoli, A**., Ghazvini, K., & Mansury, D. (2016). Prevalence of nontuberculous mycobacteria isolated from environmental samples in Iran: A meta-analysis. *J Res Med Sci, 21*, 58. Doi:10.4103/1735-1995.187306
18) Aghdam, M. N., Abbaszadegan, M. R., **Tafazoli, A**., Aslzare, M., & Mosavi, Z. (2016). Presence of the RET Cys634Tyr mutation and Gly691Ser functional polymorphism in Iranian families with multiple endocrine neoplasia type 2A. *Hormones (Athens), 15*(1), 65-72. Doi:10.14310/horm.2002.1611
19) Shekarriz, R, Faghani, S, **Tafazoli, A**., Hashemi-Soteh, M.B., (2019). The Correlation between Phospholipase C Epsilon (*PLCE1*) Gene Polymorphisms and Risk of Gastric Adenocarcinoma in Iranian Population. Int J Hematol Oncol Stem Cell Res, Doi:10.18502/ijhoscr.v13i3.1268
20) Hashemi, B., Vali nejad, A., Ataei, Gh., **Tafazoli, A.**, Ghasemi, D., Siamy, R. (2019). Knowledge and attitude toward genetic diseases and genetic tests among pre-marriage individuals: A cross-sectional study in northern Iran, International Journal of Reproductive BioMedicine, Doi:10.18502/ijrm.v17i8.4819

International but not indexed in PubMed

2 as co-author

1. Saravani, A., HassanzadehNazarabadi, Y., **Tafazoli, A**., Bagheri, Z., Soudmand, M., Sadeghi, S., HassanzadehNazarabadi, M. (2015). Prediction of genetic disorders in consanguineous marriages, using artificial Neural network. *ICE*.
2. Hashemi, B., Aliasgharian, A., **Tafazoli, A**., Karami, H., Kowsaryan, M. (2017). Frequency of Factor V Leiden, Prothrombin G20210A and MTHFR C667T Mutations in Beta-Thalassemia Patients Compare with Normal individuals from North of Iran. *Journal of Genomics & Gene Study*. imedpub.

Persian

1 as co-author

1. AkbariKordkheyli, V., Zarpour, S., Nabipour, E., **Tafazoli, A**., Bagheri, A. (2018). An Overview on the Effects of Silibinin on Different microRNAs Expression in Cancer. *J Mazandaran Univ Med Sci*. 28(165): 213-229

Journal Reviewer

- Pharmacgenomics, ISSN: 1462-2416
- Human Genomics, ISSN: 1479-7364

Guest Editor:

- Frontiers in Genetics, ISSN: 1664-8021
  – Pharmacogenetics and Pharmacogenomics section
- Frontiers in Pharmacology, ISSN: 1663-9812
  – Pharmacogenetics and Pharmacogenomics section
- Pharmaceutics, ISSN: 1999-4923