UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Variant Effect Prediction in Pharmacogenomics: Tools, Performance, and Implementation

Lore Pellens

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Bioinformatics

SUPERVISOR :
Prof. dr. Ziv SHKEDY
Prof. dr. Pieter-Jan VOLDERS

MENTOR :
De heer Sven VAN DER MAAS

2024
2025

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Variant Effect Prediction in Pharmacogenomics: Tools, Performance, and Implementation

Lore Pellens

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Bioinformatics

SUPERVISOR :

Prof. dr. Ziv SHKEDY

Prof. dr. Pieter-Jan VOLDERS

MENTOR :

De heer Sven VAN DER MAAS

Preface

After spending five memorable years as a student at Hasselt University, I have reached the stage of submitting my thesis for my Master's degree in Statistics and Data Science: Bioinformatics. These five years were a valuable experience with invaluable learning opportunities, as well as some challenges, which have helped shape my development as an individual. Thanks to the support of many people, I was able to bring this journey to a successful end.

First, I would like to acknowledge Prof. Dr. Ziv Shkedy, my internal supervisor at Hasselt University, for his formal role in this project. I am especially grateful for my external supervisors, Prof. Dr. Ir. Pieter-Jan Volders and Ir. Sven van der Maas for their expertise, thoughtful suggestions, and encouraging words that helped shape this thesis. Their guidance from within the Jessa Hospital provided a valuable clinical perspective.

I am deeply thankful to my parents, grandparents, and sister for their unconditional support, love, and belief in me. Your encouragement kept me going, and your pride means a lot to me.

To my best friends, Bauke and Tuur — thank you for always being there. Whether it was a moment of stress, procrastination, or celebration, your friendship has been a constant source of laughter, motivation, and balance.

To my animals — especially my dog, Bolt — your quiet company, comfort, and warm presence helped me more than words can say. You've been there during long days of writing and coding, reminding me to pause, breathe, and smile.

Finally, I want to thank my entire family and all my friends — near and far — for their encouragement, interest, and support throughout this journey. Every kind word or check-in helped me cross the finish line.

Lore Pellens,
June 2025

Abstract

Background: Pharmacogenomics is shaping the future of precision medicine by studying how differences in our genes affect our response to medications. Next-Generation Sequencing technologies uncover numerous rare or novel variants, many of which lack functional or clinical interpretation. The accurate prediction of these effects remains challenging due to the lack of tools that are dedicated to predicting the effects of pharmacovariants. This thesis evaluates the performance of four widely-used variant effect predictors — AlphaMissense, CADD, PolyPhen-2, and SIFT — in predicting functional classifications of pharmacovariants. Moreover, several ensemble methods combining these prediction scores were developed and evaluated to improve pharmacogenomic classification performance.

Methodology: 1 900 variants were retrieved from the PharmVar database, with 419 variants having established functional classifications ('no function', 'decreased function', 'normal function', and 'increased function') selected for analysis. Variant effect predictions were performed using Ensembl VEP, with scores extracted and compared against PharmVar functional annotations. Ensemble methods combining scores from all four predictors were developed using multinomial logistic regression, random forest and support vector machine models.

Results: The individual variant effect prediction tools performed poorly in discriminating between functional pharmacogenomic annotations, with area under the curve values ranging from 0.51 to 0.58. The ensemble support vector machine model demonstrated superior performance, with an accuracy of 37.18%, a precision of 35.59% and a recall of 37.78%. Moreover, an automated analysis pipeline was developed using Nextflow to facilitate novel pharmacovariant analysis.

Conclusion: The findings presented in this thesis highlight the limited effectiveness of current variant effect predictors in pharmacogenomics and underscore the need for tools that are dedicated to predicting the effects of pharmacovariants. Although ensemble methods, particularly support vector machines, offer moderate improvements, further research is essential to enhance the interpretation of rare and novel pharmacovariants for clinical use.

Contents

List of Abbreviations

ADME	Absorption, Distribution, Metabolism or Excretion
ADR	Adverse Drug Reaction
APF2	Activity Prediction Framework 2
AUC	Area Under the Curve
AV	Activity Value
CADD	Combined Annotation-Dependent Depletion
CNV	Copy Number Variant
CPIC	Clinical Pharmacogenetics Implementation Consortium
CYP	cytochrome P450
DPD	Dihydropyrimidine Dehydrogenase
DPWG	Dutch Pharmacogenetics Working Group
EHR	Electronic Health Record
G6PD	Glucose-6-Phosphate Dehydrogenase
IM	Intermediate Metaboliser
indels	insertions/deletions
NGS	Next-Generation Sequencing
NM	Normal Metaboliser
nsSNP	non-synonymous Single Nucleotide Polymorphism
PD	Pharmacodynamics
PGx	Pharmacogenomics
PharmGKB	Pharmacogenomics Knowledge Base
PharmVar	Pharmacogene Variation Consortium
PK	Pharmacokinetics
PM	Poor Metaboliser
PolyPhen-2	Polymorphism Phenotyping 2
ROC	Relative Operating Characteristic
SNP	Single Nucleotide Polymorphism
SO	Sequence Ontology
SV	Structural variation
UM	Ultra-rapid Metaboliser
VCF	Variant Call Format
VEP	Variant Effect Predictor
VIP	Very Important Pharmacogene
WES	Whole Exome Sequencing

1 Introduction

Adverse drug reactions represent a major challenge in modern healthcare, contributing to increased morbidity, hospitalisations, and healthcare costs (Osanlou et al., 2022). The conventional approach to medication prescription is predicated on population averages and takes into account factors such as the patient's weight and age (T P et al., 2009). However, this approach has been shown to be inadequate in addressing the broad inter-individual variability in drug response. As a result, unintended consequences have frequently been observed, including ineffective treatment outcomes and unwanted side effects (Anunobi, 2024).

Pharmacogenomics (PGx) offers a promising solution by tailoring drug therapy to an individual's genetic makeup (Weinshilboum and Wang, 2017). The field of PGx investigates how genetic variations influence drug metabolism, efficacy, and toxicity, thereby enabling more precise, effective and safer prescribing practices (Hafidh et al., 2023). The integration of PGx into clinical decision-making is a cornerstone of personalised medicine.

The advent of Next-Generation Sequencing (NGS) has made it possible to detect a large number of genetic variants in a single analysis (Russell et al., 2020). However, most of these variants are either novel or rare, and they lack clinical evidence to support their impact on protein function (Zhou et al., 2022). The experimental methods employed to assess the functional impact of each variant are expensive. Consequently, in silico tools have recently emerged to evaluate how these variants affect protein function (Zhou and Lauschke, 2021). Despite recent advancements in PGx research, novel pharmacogenomic variants could potentially contribute to unexplained variability in treatment responses. However, understanding their impact remains challenging. The transition from population-based prescribing to genetically-informed treatment is illustrated in Figure 1, which contrasts the conventional and pharmacogenomic approaches to medication selection.

This thesis aims to evaluate four widely-used variant effect predictors (AlphaMissense, CADD, PolyPhen-2 and SIFT) and assesses their accuracy in predicting the effect of pharmacogenomic variants. Furthermore, it explores whether ensemble models that combine these prediction scores can improve classification accuracy. Additionally, an automated pipeline was developed. The pipeline streamlines this analysis process, making it easier for researchers and clinicians to interpret newly discovered variants in pharmacogenes.

This thesis begins with a comprehensive literature review, including a historical perspective on PGx, its the key principles and terminology, sequencing technologies, the types of genetic variants relevant to PGx, and the most commonly involved pharmacogenes in drug metabolism. This is followed by a review of existing variant effect prediction tools and the challenges of clinical implementation. Section 3 outlines the research objectives, which focus on evaluating the predictive performance of the selected variant effect predictors.

Section 4 describes the methodology, including data collection, variant effect prediction and ensemble classifier development. This is followed by the results where the performance of the individual variant effect predictors, the outcomes of the ensemble methods and the implementation of an automated analysis pipeline based on Nextflow are presented. Section 6 discusses the implications of the findings, the methodological limitations and the need for PGx-specific tools. Sections 7 and 8 consider ethical aspects and societal relevance. Finally, section 9 provides concluding remarks and outlines directions of future research.

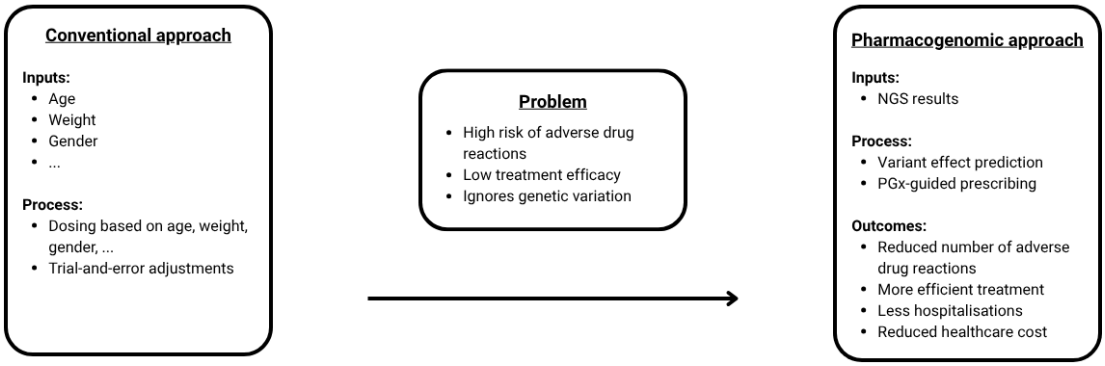

Figure 1: A conceptual overview contrasting conventional prescribing practices with PGx-guided approaches.

2 Literature review

2.1 Historical perspective

The origins of genetics date back to 510 before Christ (Pirmohamed, 2011). During this time period, the Greek philosopher and mathematician Pythagoras observed that the consumption of fava beans induced a toxic response to some people but had no harmful effect to others (Somogy, 2008). At the time, Pythagoras was unable to provide a hypothesis that would provide an explanation for the observed variation in reactions to the beans (Buguliskis, 2015). Today, the condition is referred to as favism, which is characterised by a genetic mutation in the red blood cell enzyme glucose-6-phosphate dehydrogenase (G6PD). This mutation has been observed to result in the development of potentially fatal hemolytic anemia in the presence of certain foods, drugs, or chemicals (Meletis and Konstantopoulos, 2004).

In 1959, Vogel (1959) introduced the concept of *pharmacogenetics* to describe how genetic factors influence drug responses. A series of observations indicated that the response of individuals varies greatly when exposed to equivalent doses of a pharmaceutical agent (Caraco, 2004). For instance, approximately 10% of African Americans develop hemolytic anaemia following treatment with the antimalarial drug primaquine, an adverse reaction (ADR) that is rarely observed in individuals of European ancestry (Auwerx et al., 2022). This observation is consistent with Motulsky's suggestion that, because a particular gene may be more prevalent in certain ethnic groups, any drug reaction that is more common in a particular racial group will usually have a genetic basis — provided that other environmental variables are equal (Motulsky, 1957). Decades later, the findings of Motulsky's study were confirmed by Nkhoma et al. (2009), who revealed that G6PD deficiency, due to mutations in the X-linked G6PD gene, is geographically correlated with areas inhabited by populations that have been exposed to malaria over time. G6PD deficiency tests are important in these areas, since treatment with antimalarial drugs such as primaquine can trigger hemolysis and lead to hemolytic anemia (Nkhoma et al., 2009; Mason et al., 2007).

The discovery of genetic variations in the metabolism of debrisoquine and sparteine led to notable interest in pharmacogenetics in the clinical setting (Kalow, 2006). Researchers discovered that the functional absence of the cytochrome liver enzyme CYP2D6 was responsible for both deficiencies (Owen et al., 2009). Its activity ranges from complete deficiency to overactivity, which can lead to drug toxicity or therapeutic failure at recommended drug doses (Sistonen et al., 2007).

Early pharmacogenetic research primarily investigated single-gene variants related to differences in drug metabolism (Weinshilboum and Wang, 2006). However, complex traits involving multiple genes with compensatory or overlapping roles are likely to account for most of the genetic variability in drug response (Charlab and Zhang, 2013). In contrast to pharmacogenetics, pharmacogenomics (PGx) is generally used for more advanced ap-

proaches and examines the entire genotype (genome) in relation to drug response rather than focusing only on one or a few genes (Stakos and Boudoulas, 2002). PGx holds the promise of tailoring drugs to an individual's genetic make-up, and is therefore of great interest in clinical practice (Novello et al., 2007).

2.2 Pharmacogenomic principles and terminology

Drug responses are highly variable between patients (Lam and Cavallari, 2013). The variability is largely due to genetic variations among individuals in their capacity to process and react to medications. The concept of drug response encompasses two distinct yet interconnected domains: drug disposition, also termed pharmacokinetics (PK), and drug effect, known as pharmacodynamics (PD) (Cohen and Kang, 2008). *Pharmacokinetics* is the study of how the body deals with an administered drug, including absorption, distribution, metabolism and excretion (ADME) (Eusuf and Thomas, 2019). The blood and tissue concentrations of drugs and their subsequent pharmacological or toxicological effects are determined in the ADME process (Li et al., 2019). For many drugs, efficacy, toxicity and the patient's exposure to the drug can be affected by the activity of genetic variations in PK-relevant ADME genes encoding enzymes, transporters, cell membrane and intracellular receptors or components of ion channels (Arbitrio et al., 2018). On the other hand, *pharmacodynamics* is the study of both the biochemical and physiological effects of drugs in the body, as well as the relationship between drug concentration and the effect produced by them (Neamțu, 2020). Individuals may experience treatment failure or toxicity if the variability in a drug's pharmacodynamics is not properly accounted for (Kantae et al., 2016).

An individual's response to a drug, which encompasses both positive and negative reactions, is a complex process influenced by numerous genes (Neamțu, 2020). These individual-specific responses make drug dosing challenging (Meyer et al., 2024). Patients may experience varying outcomes ranging from desired therapeutic effects to no effects or even toxicities at the standard effective dose (Tyson et al., 2020). Research indicates that only 30 to 60% of individuals treated with antidepressants, antipsychotics, β-blockers or statins respond appropriately, and that 5 to 7% of all hospital admissions are attributable to ADRs (van der Drift et al., 2023). This underscores the challenges of optimising drug therapy.

Cytochrome P450 enzymes (CYPs) are of particular importance in PGx because of their central role in drug metabolism (Lonah et al., 2023). Their primary function is to metabolise a wide variety of xenobiotics and clear potentially toxic compounds from the body (Stocco and Tyndale, 2022). This broad group of xenobiotics includes drugs, environmental pollutants, cosmetics and food additives (Esteves et al., 2021). While generally harmless, they can be potentially toxic. Drug developers and researchers study how different drugs can affect, or be affected by, the activity of CYP enzymes, which can lead to unexpected clinical outcomes (Gilani and Cassagnol, 2025). Phenotypic changes in metabolising activity are typically classified in four groups: poor metabolisers (PMs), intermediate metabolis-

ers (IMs), normal metabolisers (NMs) and ultra-rapid metabolisers (UMs), all four are attributable to drug response due to genetic variations in CYP genes (Zhao et al., 2021).

One of the most important CYP enzymes is CYP2D6 (Taylor et al., 2020). It is a cytochrome P450 enzyme encoded by the *CYP2D6* gene and plays a crucial role in the metabolism of approximately 25% of commonly prescribed drugs (Rüdesheim et al., 2022; Berg et al., 2021). Genetic variations in the *CYP2D6* gene influence the enzyme's activity, affecting drug response among individuals (Stojanović Marković et al., 2022). In 1996, Daly et al. proposed a method for categorising the allelic variants of *CYP2D6*, the star-allele nomenclature. In the star-allele nomenclature, *1 is the reference sequence against which polymorphic sites are compared (Robarge et al., 2007). This is usually the first sequence described that encodes a functional protein product. When a new variant is identified with a nucleotide change that results in an amino acid substitution or is shown to affect transcription, splicing, translation or post-transcriptional or post-translational modification, then a unique number (e.g. *CYP2D6**3) is assigned (Lee et al., 2019a). Non-functional nucleotide changes that are thought to occur on the same chromosome or to be inherited with a named star allele are defined by an additional number (e.g., *CYP2D6**2.002, *CYP2D6**2.003) (Gaedigk et al., 2019). The main star allele is denoted by the additional number 001 (e.g., *CYP2D6**3.001). Where multiple variant alleles on the same chromosome are shown to have a functional effect on the protein in a context where no single polymorphism has an effect, a new allele number is assigned (e.g. *CYP2D6**17). All alleles with the same star number, also referred to as suballeles, are assumed to have an equivalent function (Nofziger et al., 2020).

To facilitate the interpretation of the activity of the CYP2D6 enzyme, Gaedigk et al. (2008) developed the CYP2D6 Activity Score System. Based on the function assigned to the group of alleles with the same star number to which the particular star allele belongs, the allele receives an activity value (AV) ranging from zero to one (e.g., zero for no function, 0.5 for decreased function and one for normal function) (Caudle et al., 2020). If an allele contains multiple copies of a functional gene, the value is multiplied by the number of copies present (Crews et al., 2021). Every person inherits one paternal and one maternal allele, forming the person's diplotype (Shin et al., 2019). The sum of the activity values of each allele in the diplotype equals the activity score of the diplotype (Caudle et al., 2020). The activity score can then be translated into the associated phenotype (Gaedigk et al., 2017). Figure 2 provides a visual representation of the processes from star allele to activity score and from diplotype to phenotype.

The Activity Score axis in Figure 3, represents the range of values assigned to each type of CYP2D6 metaboliser. This figure also provides an overview of the implications and recommendations for codeine therapy based on the *CYP2D6* phenotype. CYP2D6 converts codeine to its active metabolite, morphine, which is responsible for its pain-relieving effect (Dean and Kane, 2021).

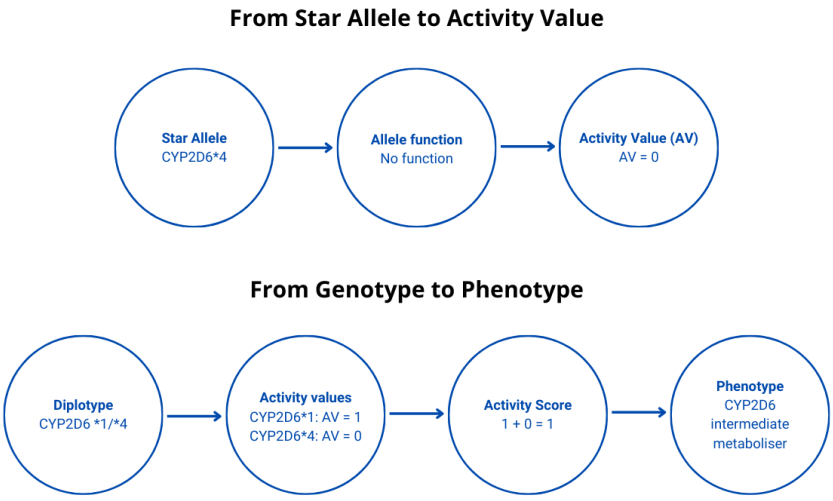

Figure 2: Process from star allele to activity value and from genotype to phenotype.

CYP2D6 PMs carry two defective alleles and therefore produce almost no morphine after administration of codeine (Wollmann et al., 2023). The activity score of the PMs is equal to zero (Crews et al., 2021). A higher risk of adverse drug reactions may be experienced by individuals with PMs when they are treated with drugs in which the CYP2D6 enzyme plays a key role in the process of drug deactivation (Magarbeh et al., 2021). However, in the context of prodrugs such as codeine, these individuals may exhibit a reduced therapeutic response. The Dutch Pharmacogenetics Working Group (DPWG) advises that in cases of PMs, the use of an alternative painkilling agent should be considered (Matic et al., 2022).

Individuals classified as CYP2D6 IMs exhibit reduced, though not absent, enzyme activity leading to decreased conversion of codeine to morphine (Gaedigk et al., 2017; Dean and Kane, 2021). Their CYP2D6 activity scores range from greater than zero to less than 1.25 (Crews et al., 2021). To ensure an adequate analgesic effect, the DPWG recommends either increasing the dose of codeine or using an alternative analgesic (Matic et al., 2022).

NMs, on the other hand, show expected CYP2D6 activity and corresponding morphine formation following codeine administration (Carranza-Leon et al., 2021). Their activity score fall between 1.25 and 2.25 (Caudle et al., 2020). In these individuals, codeine can generally be used safely, provided that the dose is adjusted according to the patient's age and weight (Crews et al., 2021).

UMs have markedly increased CYP2D6 activity, reflected by activity scores greater than 2.25 (Magarbeh et al., 2021; Dean and Kane, 2021). The increased enzymatic activity leads

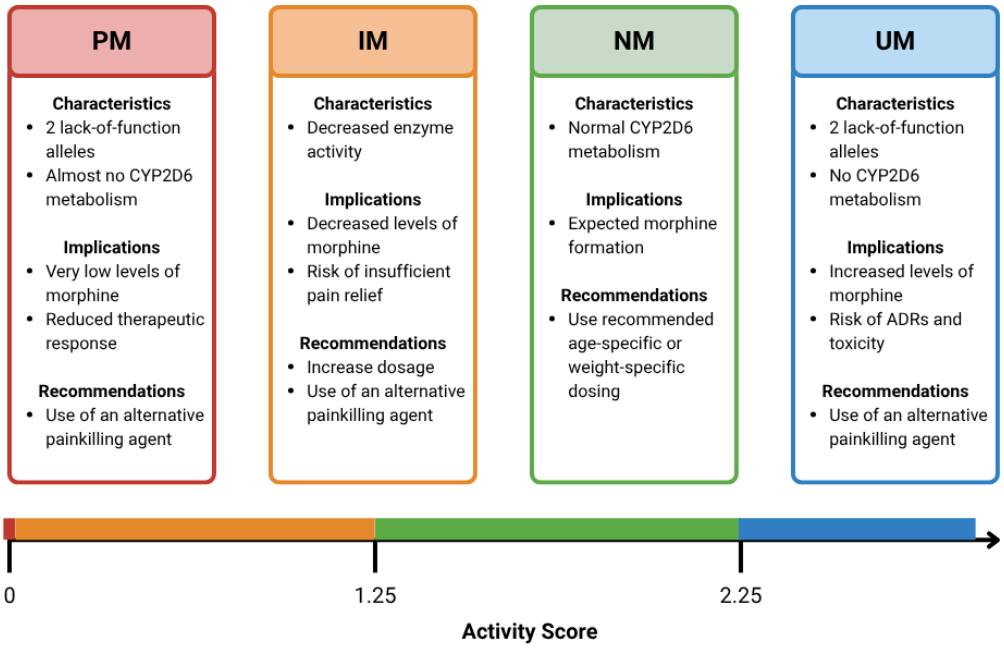

Figure 3: Overview of the metabolism of codeine by the CYP2D6 enzyme based on the four metaboliser types: poor metabolisers (PM), intermediate metabolisers (IM), normal metabolisers (NM) and ultra-rapid metabolisers (UM).

to rapid and excessive conversion of codeine to morphine, consequently raising the risk of serious or even life-threatening ADRs. Given the risk of morphine toxicity, avoidance of codeine and use of an alternative painkilling agent is recommended (Crews et al., 2021).

A comprehensive understanding of the molecular mechanisms underlying a drug's action on these enzymes is thus essential to ensure that patients receive appropriate therapy.
Beyond the CYP family, various other enzyme systems contribute to pharmacogenomic variability. For instance, the thiopurine S-methyltransferase (TPMT) enzyme, which is responsible for the metabolism of thiopurine drugs, exhibits genetic polymorphisms that result in variable enzyme activity (Zhou et al., 2020). The frequency of individuals carrying two non-functional TPMT alleles varies by ethnicity, ranging from 0.03% to 0.56% (Zhou and Lauschke, 2022b). Having two non-functional alleles remarkably increases the risk of life-threatening myelosuppression when given standard doses of thiopurine drugs.

2.3 Sequencing technologies in pharmacogenomics

The assessment of genetic variants is a key component of PGx (van der Lee et al., 2020). Since its development by Sanger et al. (1977), Sanger sequencing has been considered the gold standard for identifying nucleotide sequence variations. Next-generation sequencing (NGS) has transformed genetic research by offering high-throughput, cost-effective analysis of DNA. Depending on the application, NGS can be applied at different scales - from targeted panels to full exome or genome sequencing - providing flexibility for both clinical and research settings (Schwarz et al., 2019). Currently, NGS technologies are widely used in PGx research and clinical practice (Tafazoli et al., 2021). This is because NGS technology enables parallel sequencing (Gerilovych et al., 2024). Parallel sequencing technologies have revolutionised genomics by significantly increasing the speed, throughput and cost-effectiveness of DNA sequencing (Abdi et al., 2024). The following paragraphs describe three main types of NGS - targeted gene panels, exome-wide NGS and genome-wide NGS.

2.3.1 Targeted gene panels

Targeted gene panels analyse a predefined set of genes, typically those associated with a specific phenotype (Shah et al., 2020). To investigate gene-drug associations, laboratories often select panels that target genes involved in known drug pathways (Ji and Shaaban, 2024). By limiting the number of genes sequenced, these panels reduce the cost of achieving adequate coverage by maximising sequencing efficiency and minimising computational and storage requirements (Rehder et al., 2021). Additionally, because they focus mainly on known drug pathways, targeted gene panels are generally not designed to detect novel variants (Enko et al., 2023).

2.3.2 Exome-wide next-generation sequencing

Exome-wide NGS, also known as Whole Exome Sequencing (WES), is focused on capturing and sequencing the protein-coding regions of the genome (Satam et al., 2023). These regions are collectively referred to as the exome. The exome accounts for 1.5% of the entire genome (Jelin and Vora, 2018). WES allows for a comprehensive analysis of both common and rare genetic variants associated with drug treatment outcomes (Wang et al., 2024). Although WES offers broader coverage than targeted gene panels, it does not capture the non-coding part of the genome, which plays an important role in gene regulation and protein folding (Burdick et al., 2020). If this part of the genome is also of interest, WES is less suitable.

2.3.3 Genome-wide next-generation sequencing

Genome-wide NGS, otherwise referred to as Whole Genome Sequencing (WGS), is a technique that provides an overview of the entire human genome, including non-coding regions (Zhou et al., 2022). One of the main differences between WGS and other forms of

NGS is the considerably larger volume of data it generates. (Bagger et al., 2024). A key application of WGS is the discovery of genetic variants and their association with known and previously uncharacterised clinical conditions (Austin-Tse et al., 2022). Compared to WES, the application of WGS has been demonstrated to generate a substantially higher number of variants (Warr et al., 2015). This is attributable not only to the larger sequencing space, but also to the fact that regions outside the exome are less evolutionarily conserved.

Mizzi et al. (2014) employed WGS to identify novel, potentially clinically relevant variants affecting the structure and function of 231 pharmacogenes across 481 human genomes representing diverse ethnic groups. This study aimed to investigate the advantages of this approach over conventional genetic screening methods. The study demonstrated that WGS can reveal a significant number of unique or rare pharmacogenomic markers that would otherwise remain undetected by conventional methods, such as PCR or microarray-based methods.

2.4 Genetic variants in pharmacogenomics

Recent advancements in sequencing technologies have facilitated the discovery of genetic variation within the human genome (Russell et al., 2020). Genetic factors are among the most important contributors to inter-individual variability in drug response (Katara and Yadav, 2019). Understanding the relationship between genetic variations and drug response is essential for optimising pharmacotherapy (Russell et al., 2020). However, sequencing techniques typically sequence DNA samples containing both maternal and paternal DNA, without distinguishing which variants originate from which parent (Choi et al., 2018). In other words, it does not take into account the *phase* of the DNA in these samples, i.e. the specific arrangement of variants on each of the two homologous chromosomes, collectively referred to as the *diplotype*.

The necessity of haplotype phasing is illustrated by Figure 4, with *CYP2B6* serving as an example. The inability to phase the rs3745274 (NC_000019.10:g.41006936G>T) and rs2279343 (NC_000019.10:g.41009358A>G) variants to the correct allele can result in differences in haplotype assignment (van der Lee et al., 2020). The most common situation in individuals who are heterozygous for both variants is shown on the right side of the figure, the CYP2B6*1/*6 diplotype. The left side shows an alternative configuration where the variants are located on opposing alleles, resulting in a CYP2B6*4/*9 diplotype. Consequently, it is not possible to determine directly on which parental chromosome a given allele resides (Browning and Browning, 2011).

Therefore, only by performing phasing or haplotyping across the entire gene region, the exact functional effect of an allele can be determined (Hari et al., 2023). In PGx, where most known pharmacogenetic haplotypes are defined by specific combinations of genetic variants, these haplotypes are often classified using the star-allele nomenclature as described

in section 2.2.

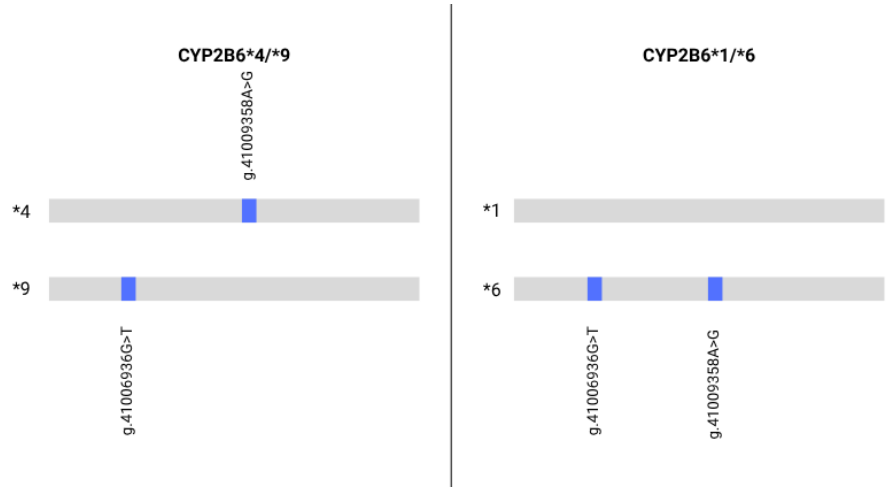

Figure 4: An illustrative example of the principle of haplotype phasing for CYP2B6.

2.4.1 Types of variants relevant to pharmacogenomics

The effectiveness of drugs varies greatly between individuals (Davis and Limdi, 2021). In fact, pharmacological treatment is unsuccessful for 40-70% of patients. This means these patients either experience ADRs or they demonstrate a lack of efficacy. It has been estimated that 15–30% of this observed variability can be accounted for by genetic polymorphisms (Zhou et al., 2017). Genetic polymorphism is defined as having two or more alternative forms of an allele in an individual's genome, resulting in varying phenotypes within the same population (Sameer et al., 2021). Single nucleotide polymorphisms (SNPs), copy number variants (CNVs), insertions or deletions (indels) and structural variants (SVs) are the four most common types of polymorphism (Srinivasan et al., 2016). Although most such polymorphisms are rare and have low allele frequencies, pharmacogenetic testing in clinical practice is currently limited to validated and experimentally characterised variants that allow qualified predictions to be made about their phenotypic consequences (Zhou et al., 2017).

The 1000 Genomes Project has sampled and sequenced approximately 3,200 individuals from 26 populations in Africa, East Asia, Europe, South Asia and the United States using both WGS and targeted exome sequencing (Gustafson et al., 2024; Belsare et al., 2019). The project identified over 88 million genetic variations, including 84.7 million SNPs, 3.6 million short indels and 60,000 SVs. The project's findings on the prevalence of variation are of great value (Birney and Soranzo, 2015). Furthermore, it provides an understanding of how genetic variation can differ between people from different continents, advancing

knowledge of recent human evolution and medicine.

The star allele nomenclature proposed by Daly et al. (1996) was introduced in section 2.2 and was illustrated using the *CYP2D6* gene. This nomenclature was first used to identify alleles within the cytochrome P450 (CYP) gene family (Robarge et al., 2007). The nomenclature was later extended to almost all genes studied in PGx. Each star allele corresponds to a specific haplotype and is often linked to a functional phenotype (normal, decreased, increased or no enzymatic function) (Lee et al., 2019b). This translates to the four metaboliser types described in section 2.2.

2.4.2 Pharmacogenes and their role in drug metabolism

Some genetic variants, such as those in the cytochrome P450 (CYP) genes, affect several drugs in different classes, while other gene-drug pairs are more specific (White et al., 2022). These genes, which have been shown to influence drug response, are commonly referred to as *pharmacogenes* or PGx genes (Katara and Yadav, 2019). Zhou and Lauschke (2022a) conducted a study incorporating WES and WGS data from 141,614 unrelated individuals across 12 human populations with the aim of extending current knowledge of the genetic landscape of major drug-metabolising CYP genes. The study revealed that uncharacterised rare alleles account for between 1.5% and 17.5% of the total genetically encoded functional variability, highlighting the influence of common and rare genetic variations on outcomes of pharmacotherapy.

Figure 5 provides an overview of nine pharmacogenes and the types of drugs whose metabolism they influence. It should be noted that this figure only provides a few examples of gene-drug relatedness. The genes may also be involved in the metabolism of other types of drugs. The nine pharmacogenes are classified as Very Important Pharmacogenes (VIPs) by the Pharmacogenomics Knowledge Base (PharmGKB). The Pharmacogene Variation Consortium (PharmVar) provides a repository and the nomenclature of genes which contribute to the variability in drug metabolism and response, including the nine genes presented in Figure 5 (Gaedigk et al., 2021). Moreover, the Clinical Pharmacogenetics Implementation Consortium (CPIC) provides guidelines for translating the genotype information of the nine VIPs into actionable prescribing recommendations (Relling and Klein, 2011).

As illustrated in Figure 5, *CYP2D6*, *CYP2C19* and *CYP2B6* are involved in the metabolism of antidepressants (Bousman et al., 2023). Patients may be at risk for poor therapeutic outcomes because they have *CYP2D6*, *CYP2C19*, or *CYP2B6* allelic variants that alter the biotransformation of antidepressants. In addition to its role in the metabolism of antidepressants, *CYP2D6* is also involved in the metabolism of anti-cancer agents, such as tamoxifen. (Taylor et al., 2020). Tamoxifen inhibits tumour growth and promotes apoptosis in oestrogen receptor-positive tumours (Mulder et al., 2021). This reduces the risk of breast

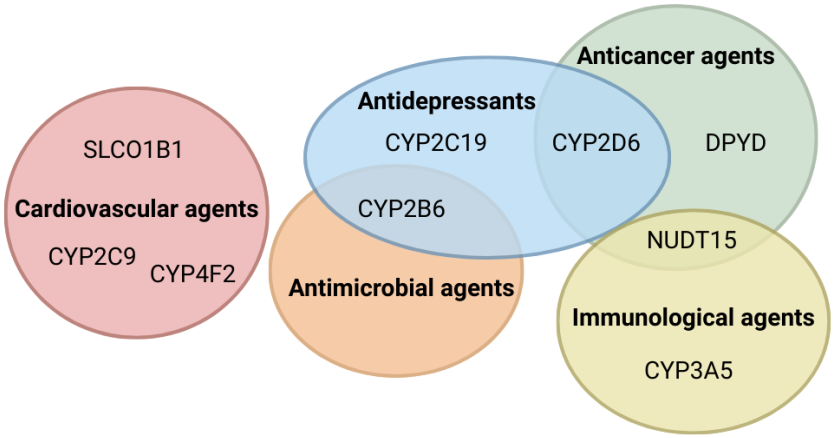

Figure 5: Venn diagram of nine pharmacogenes listed by PharmVar and provided with CPIC guidelines. The genes are grouped based on the drug class they are involved in metabolising.

cancer recurrence and death. It has been estimated that the proportion of individuals with poor or ultrarapid CYP2D6 metabolism is 5.4% and 3.1% in Europe, 1.9% and 4.6% in the Americas, and 0.4% and 21.2% in Oceania, respectively. Therefore, the impact of these metabolic phenotypes on tamoxifen treatment is unlikely to be negligible, potentially resulting in reduced efficacy for hundreds of thousands of breast cancer patients (He et al., 2020). Marcath et al. (2017) used patients' genetics to adjust the dose of tamoxifen for those with low-activity *CYP2D6* genotypes. This approach increases treatment efficacy without increasing toxicity related to the treatment. A second class of medical agents impacted by *CYP2B6* are antimicrobial agents such as artemisinin derivatives (Langmia et al., 2021). While UM phenotypes are associated with reduced drug exposure, PM variants are linked to increased plasma concentrations of artemisinin (Soyinka et al., 2022). However, no studies have reported an association between *CYP2B6* gene variants and the efficacy of artemisinin.

Another clinically important pharmacogene is *CYP3A5*, which metabolises tacrolimus, an immunosuppressant used to prevent organ rejection in transplant recipients (Barbarino et al., 2013). Individuals with at least one functional allele generally have lower trough concentrations of tacrolimus than individuals with two non-functional alleles (Birdwell et al., 2015). This can result in a delayed achievement of the target blood concentration.

Mercaptopurine and azathioprine are generally used to treat non-malignant immunological disorders, whereas thioguanine is used to treat leukemia (Relling et al., 2019). All three of these medical agents belong to the thiopurine class of drugs and are metabolised by the *NUDT15* gene.

Like thioguanine, fluoropyrimidines are anti-cancer drugs (Amstutz et al., 2018). Toxicity associated with these drugs is often caused by reduced activity of the enzyme dihydropyrimidine dehydrogenase (DPD), which is the main enzyme responsible for the inactivation of fluoropyrimidines (Henricks et al., 2018). The *DPYD* gene encodes DPD.

The anticoagulant warfarin is metabolised by both *CYP4F2* and *CYP2C9* (Johnson et al., 2017). Warfarin is a commonly prescribed blood thinner to prevent blood clots in patients with deep vein thrombosis, atrial cardiac arrhythmias, or prosthetic heart valves (Lee and Klein, 2013). Another pharmacogene associated with the metabolism of a cardiovascular agent is *SLCO1B1* (Cooper-DeHoff et al., 2022). *SLCO1B1* encodes a transporter that facilitates the uptake of statins in the liver. Statins are medications with powerful cholesterol-lowering properties and have made outstanding contributions to the prevention of cardiovascular disease (Sirtori, 2014).

2.5 Variant Effect Prediction Tools

Sequencing provides the opportunity to detect both common and rare variants, but uncovering the functional effects of rare variants, particularly with respect to drug response, lack of response, or ADR, remains a challenge (Russell et al., 2020). Variant Effect Predictors (VEPs) can be used. VEPs are software tools that accept (human) genetic variants and predict the functional effects of these variants (Riccio et al., 2024). In order to interpret variants, certain requirements must be met (Hunt et al., 2022). This involves mapping the variants to transcripts and predicting molecular consequences. Variant interpretation requires integration of all available knowledge, and predictive algorithms are used to evaluate the impact of changes to a certain locus. Most prediction algorithms have been developed to identify disease-causing variants, making them less suitable for PGx implementation (Zhou and Lauschke, 2021). The general workflow of a variant effect predictor is illustrated in Figure 6.

Most variant effect predictors use Variant Call Format (VCF) as input data, since this is the standard exchange format used in next-generation sequencing pipelines (McLaren et al., 2016). A reference genome is required in order to map the input data (Cunningham et al., 2022). This makes it possible to determine the genomic context of the variants, as well as any differences and similarities to the reference. The next important step is transcript annotations, as illustrated in Figure 6. Since a single gene can have multiple transcripts due to alternative splicing, the functional impact of variants on the protein can differ depending on which transcript is affected and in which tissue it is expressed (Koroglu and Bilguvar, 2025). Therefore, accurately interpreting the variants requires annotating their effects in every transcript. The annotation process is followed by the prediction of the functional impact of the variants. Variant effect predictors are typically specialised in identifying one or a few categories of variants, such as SNVs, indels, missense variants or SVs (Riccio

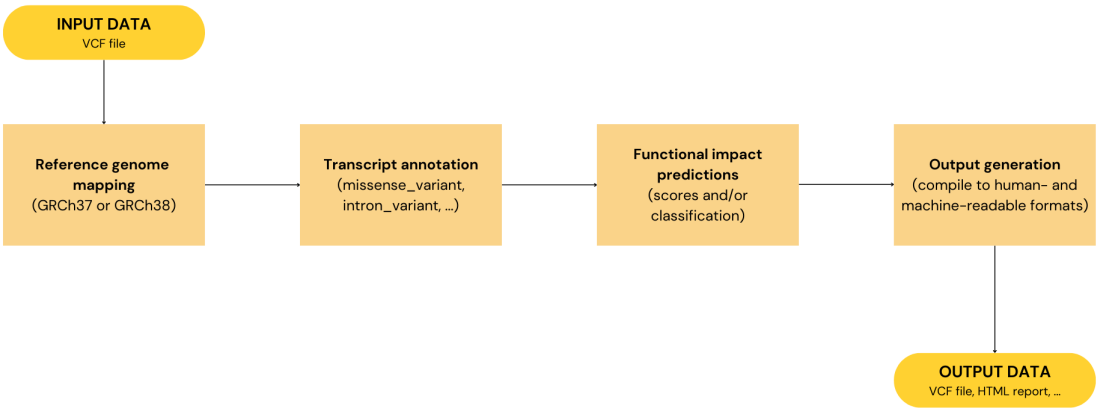

Figure 6: The general workflow of variant effect predictors.

et al., 2024). Regarding the features they use in prediction, the predictors can typically be categorised as homology sequence based or structural based models (Liu et al., 2022). Predictors based on sequence homology make the assumption that amino acid changes in conserved sequences are more likely to be deleterious when identified by searching homologous sequences across species than in other non-homologous positions, which are considered to be tolerant (Cooper and Shendure, 2011). Structural-based predictors take protein structure and folding into account to determine the effect of the associated variant (Gerasimavicius et al., 2025). Depending on the prediction tool used, either scores or probabilities are calculated, or classification is employed to assign variants to pathogenicity classes (Wang et al., 2022). The workflow ends with the generation of output in human- and machine-readable formats. Examples of output data include VCF files and HTML summary reports (McLaren et al., 2016). Several variant effect predictors will be introduced in the following sections.

2.5.1 Ensembl VEP

Ensembl is a freely available platform that provides open-source bioinformatics tools (Schubach et al., 2024). For more than 20 years, Ensembl has been developing systems to provide reference genome assemblies from public archives for interpreting genes, regulatory regions and comparative data (Harrison et al., 2024). The Ensembl Variant Effect Predictor (Ensembl VEP) is an Ensemble tool and provides methods for taking a systematic approach to predict the functional impact of genetic variants (McLaren et al., 2016).

An increasing number of scoring algorithms are being developed to aid variant interpretation (Hunt et al., 2022). For this reason, Ensembl VEP contains pre-calculated scores for over 20 of these algorithms including CADD, PolyPhen, AlphaMissense and SIFT. Since the predictor scores are pre-calculated, it ensures fast results (McLaren et al., 2016). The

disadvantage of this is that, for some tools, only variants known to the Ensembl VEP can be analysed.

2.5.2 AlphaMissense

AlphaFold is a machine learning approach developed by DeepMind that incorporates physical and biological knowledge about protein structure into the design of the deep learning algorithm, leveraging multi-sequence alignments (Jumper et al., 2021; Varadi et al., 2024). AlphaMissense is an adaptation of AlphaFold that has been fine-tuned using databases of human and primate variant population frequencies to predict the pathogenicity of missense variants (Cheng et al., 2023). AlphaMissense addopts two key components of AlphaFold: a highly accurate model of protein structure and the ability to acquire knowledge about evolutionary constraints from related sequences.

AlphaMissense undergoes training in two stages. The first stage is similar to AlphaFold in that it trains the network to predict the 3D structure of a reference sequence (Varadi et al., 2024). In the second stage, the model is trained on human proteins so that the pathogenicity of variants of the reference sequence can be predicted. Scores for AlphaMissense range from zero to one and are calibrated using a ClinVar evaluation set (Cheng et al., 2023). They can be interpreted as the approximate probability of a variant being clinically pathogenic.

In the study by Zhou et al. (2024), the performance of several predictors, including Alphamissense, was assessed on pharmacovariants. The constructed data set comprised all missense variants with phenotypic annotations by CPIC and variants with high-quality experimental characterisation data found in the literature. For the entire set variants a definition is provided regarding the categorisation of variants as either deleterious or functionally neutral. AlphaMissense demonstrated an excellent level of specificity (94%) but a relatively low level of sensitivity (33%).

2.5.3 Combined Annotation-Dependent Depletion (CADD)

Combined Annotation-Dependent Depletion (CADD) is a variant effect predictor for objectively combining many different annotations into a single measure, the C-score, for any possible human SNV or small indel event (Kircher et al., 2014). CADD, unlike other machine learning ensemble methods described in this section, not only focuses on genetic variants impacting health and disease states. Instead, CADD assumes that most of the variants that remain in humans after millions of years of natural selection are harmless or neutral (Rentzsch et al., 2019). These variants are called 'proxy-neutral'. The proxy-neutral variants are compared with a set of simulated variants that have not been purified by selection. Many of these variants are neutral, however some of them would be harmful, these variants are called 'proxy deleterious'.

Logistic regression models form the basis of CADD (Schubach et al., 2024). The raw scores are those obtained directly from logistic regression. These scores reflect the degree to which the variant is likely to have derived from the proxy-neutral (negative values) or proxy-deleterious (positive values) class.

To improve interpretability, the raw scores are converted into PHRED-like rank scores, which are based on the genome-wide distribution of scores for all potential SNVs (Rentzsch et al., 2019; Ewing and Green, 1998). For example, regardless of the details of the annotation set, model parameters, etc., a scaled score of 10 indicates a raw score in the 90th percentile of all possible reference genome SNVs, while a score of 20 or greater indicates a raw score in the 99th percentile%.

The performance of CADD was assessed by Mahmood et al. (2017). Seven different data sets were used, each comprising variants classified as deleterious or benign. The Area Under the Curve was used as a metric to assess the performance of CADD. This value exhibited a range between 0.556 and 0.939 depending on the data set. This considerable variability raises concerns regarding the practical application of CADD.

2.5.4 Polymorphism Phenotyping 2 (PolyPhen-2)

Polymorphism Phenotyping (PolyPhen-2) is a tool dedicated to the automated functional annotation of coding non-synonymous SNPs (nsSNPs), i.e. SNPs located in coding regions that result in amino acid variation in the protein products of genes (Ramensky, 2002). The prediction is based on a set of sequence, phylogenetic, and structural features that characterise the amino acid change (Adzhubei et al., 2013). For the given amino acid substitution in the protein, PolyPhen-2 passes the features to a probabilistic classifier. The output score is the probability of the substitution being damaging.

2.5.5 Sorting Intolerant From Tolerant (SIFT)

A protein may be able to tolerate an amino acid change and still function normally, or it may be intolerant to the amino acid change (Vaser et al., 2016). The Sorting Intolerant From Tolerant (SIFT) tool classifies an amino acid change as tolerated or deleterious to protein function. SIFT is a tool based on sequence homology for sorting intolerant and tolerant amino acid substitutions and predicting whether an amino acid substitution at a particular position in a protein has a phenotypic impact (Ng and Henikoff, 2001). SIFT is based on the evolutionary conservation of amino acids within protein families (Kumar et al., 2009). Positions that are highly conserved tend to be intolerant of substitution, while those with a low degree of conservation are tolerant of most substitutions.

SIFT outputs a score between 0 and 1 with a cutoff of 0.05 (Hassan et al., 2019). Amino acid substitutions with a score less than 0.05 are predicted to be deleterious and those with a score equal to or greater than 0.05 are predicted to be tolerated.

2.6 Clinical practice and challenges

In some healthcare systems, electronic health records (EHRs) are used to help clinicians integrate PGx into their practice (Caraballo et al., 2019). If the EHR contains genetic information, the computerised clinical decision support (CDS) rules alert the prescriber to a potential drug-gene interaction for certain medications (Caraballo et al., 2017). These drug-gene alerts are similar to the alerts for drug-drug interactions that are already in place at many institutions and pharmacies (Nicholson et al., 2021). Practitioners may also have access to a computerised resource with additional pharmacogenomic information to supplement the brief details provided in an alert. This shows that the implementation of PGx is gradually increasing.

Several studies provide robust evidence of the benefits of PGx-guided therapeutic strategies (Kabbani et al., 2023). For this reason, considerable effort has been directed towards implementing PGx in routine clinical practice. Successful implementation of PGx testing in this way assists both patients and providers in making therapy decisions (Haga and LaPointe, 2013). Some pharmacogenomic tests are already being used in clinical settings, demonstrating how integrating pharmacogenomic data into routine care can improve patient safety in a cost-effective way (Peruzzi et al., 2025). Despite growing enthusiasm and evidence of successful PGx testing implementation efforts, a consensus on the best way to integrate PGx testing into clinical practice has not yet been reached, and numerous challenges to its broader adoption remain (Maruf and Bousman, 2022).

Despite being generally aware of the importance of PGx and having a positive attitude towards its ability to improve drug therapy and reduce adverse effects, survey data have consistently shown that relatively few healthcare providers have adopted PGx testing (Kabbani et al., 2023). This is primarily attributed to insufficient education, which may hinder the clinical implementation of PGx testing (Luzum et al., 2021). PGx education should focus on developing knowledge and skills, such as interpreting test results and contextualising them when making treatment decisions (Just et al., 2019). In addition, education can help to close the existing gap between PGx treatment guidelines and clinical practice.

Cost is also considered to be an important factor in the application of PGx in clinical practice by healthcare systems and patients, and it is often ranked as a major barrier to implementation (Morris et al., 2022). Third-party payers, especially those who provide health insurance, want to know if PGx test coverage will lead to a decrease or increase in healthcare costs in the future. Another important insight regarding the costs of implementing PGx is that many genetic variants have low population allele frequencies (Pirmohamed, 2023). This means that trials with large sample sizes would be needed, but these might not be feasible because of both cost and difficulties in obtaining participants. It should be noted that the implementation of PGx in clinical care has a promising future, given the expectation that whole genome sequencing (WGS) will be cost-effective and technically

feasible for all clinical testing procedures (Ji and Shaaban, 2024).

PGx requires robust ethical and legal frameworks, as human DNA is a highly sensitive area (Shoaib et al., 2017). For instance, pharmacogenomic screening has consequences not only for the individual being analysed, but also for relatives (Cai et al., 2020). Information that is unanticipated is likely to arise when performing methods that evaluate either whole genome sequencing information or large panels of pharmacogenomic variants (Williams and Schoonmaker, 2023). For instance, a case of non-paternity could be revealed in a family undergoing treatment for an inherited cancer, since family members may share both pharmacogenomic variants and genetic markers of disease.

Despite the challenges that lie ahead, the potential benefits of PGx remain a strong motivation to overcome the problems preventing its adoption in clinical practice. Addressing these challenges and translating pharmacogenomic discoveries into concrete improvements in patient care requires collaborative efforts involving researchers, clinicians, regulatory bodies and industry stakeholders.

3 Research objectives

A large number of variants can be identified by Next-Generation Sequencing (NGS) technologies, most of which are either novel, rare or lack clinical evidence regarding their impact on protein function (Pandi et al., 2021). Evaluating the functional impact of a multitude of variants through functional expression assays can be costly. As an alternative, in silico variant effect predictors (VEPs) have been developed to estimate the potential impact of variants on protein function (AlSaeed et al., 2024). However, the utility of these tools in pharmacogenomics (PGx) is limited. Most VEPs are trained on datasets of disease-causing variants, where variants are classified as pathogenic or benign (Zhou et al., 2018). In PGx, the classification paradigm differs, focusing on how variants affect drug metabolism, typically categorised into metaboliser phenotypes: poor, intermediate, normal, rapid, and ultra-rapid metabolisers (Zanger et al., 2004). Predictors optimised for pathogenicity may therefore underperform when applied to pharmacovariants (Tremmel et al., 2025). This hinders their effectiveness in PGx-guided clinical decision-making, for example when making treatment recommendations based on CPIC guidelines (Relling and Klein, 2011).

The aim of this thesis is to evaluate the performance of variant effect predictors on known pharmacogenomic variants. The specific objectives are:

1. Assessment of the performance of four widely used VEPs (AlphaMissense (Cheng et al., 2023), CADD (Kircher et al., 2014), PolyPhen-2 (Ramensky, 2002) and SIFT (Ng and Henikoff, 2001)) on a curated set of known variants.

2. Exploration of ensemble methods by combining predictions from the individual VEPs using machine learning approaches, with the goal of improving predictive accuracy in the context of PGx.

3. Development of an automated analysis pipeline that facilitates future pharmacogenomic analysis and integration in clinical care.

By addressing these objectives, this thesis aims to inform the development of more PGx-specific predictive models and contribute to the optimisation of pharmacogenomic variant interpretation.

4 Methods

4.1 Data description

To ensure comprehensive and up-to-date coverage of known pharmacogenomic haplotypes, the human variants were obtained from the PharmVar API. All star alleles available in the PharmVar database (version 6.2.2) were initially retrieved, after which the variants of these alleles were selected. To ensure compatibility with modern bioinformatics tools and annotations, only variants mapped to the GRCh38 human genome assembly were included. This was necessary because variant effect predictors, such as Ensembl VEP, require genomic coordinates aligned with the reference genome used in their annotation databases.

The selected variants were formatted in the standard Variant Call Format (VCF version 4.2), with each entry detailing the chromosome number, position, rs identifier (if present), reference allele and alternate allele. Additionally, the CPIC clinical function annotations were retrieved separately from the PharmVar database to facilitate supervised learning during the ensemble method development. The final VCF file included a total of 1,900 variants.

Figure 7 summarises the number of variants for each gene included in this study. Notably, the *DPYD*, *CYP2D6* and *CYP2A6* genes have the highest variant counts, which together accounted for 68.47% of the total variants. The variants were also categorised according to their predicted molecular consequence defined by the Sequence Ontology (SO) (Eilbeck et al., 2005). Missense variants, which result in single amino acid substitutions, comprised the majority of variants in this study (70.37%).

Table 1 provides an overview of the number of variants assigned to each PharmVar function. In this study, variants with established functional classification ('no function', 'decreased function', 'normal function' and 'increased function') by PharmVar were included. A substantial proportion of variants in PharmVar lack definitive functional ('unknown or uncertain' and 'no function assigned') annotations (n = 1481), indicating that there are still

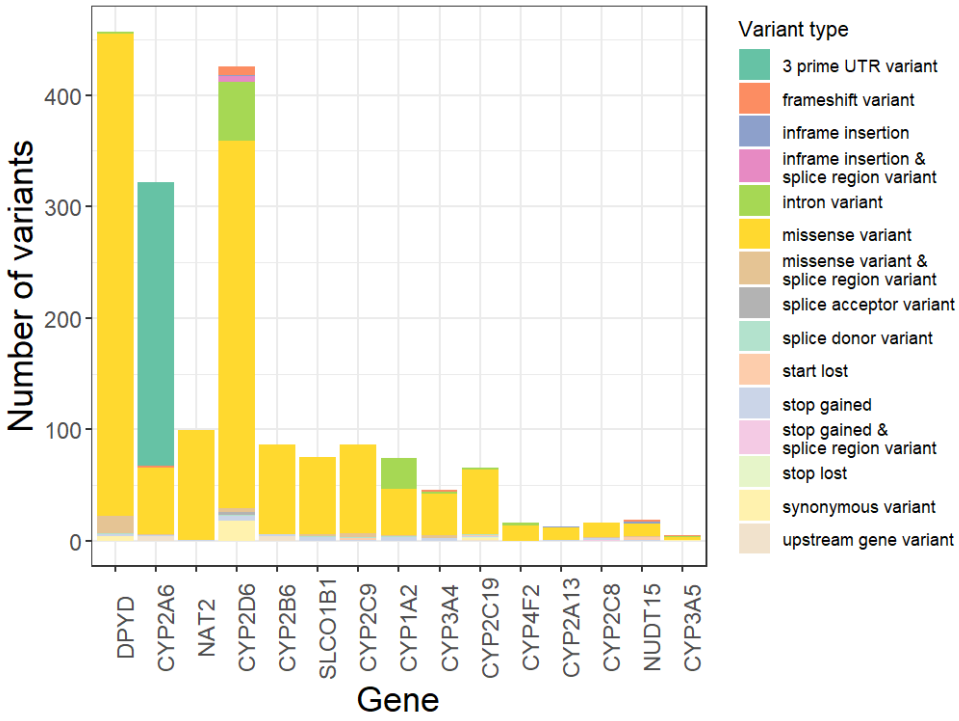

Figure 7: The number of variants that belong to each gene coloured by the variant type.

many gaps in the understanding of variant functionality. This also highlights the need for accurate variant effect predictors. When comparing variant effect predictions with the functional impact assigned by PharmVar, these four classes will be considered, resulting in a final dataset of 419 variants. Furthermore, it is important to note that these four classes are highly imbalanced: only seven variants are present with increased function. At the other extreme, the 'no function' variant class contains 189 variants. The groups with decreased and normal functionality lie between these extremes, with 127 and 96 variants, respectively. This should be considered when developing the ensemble method as described in section 4.3.

4.2 Variant effect predictors

This thesis evaluates four widely used variant effect predictors: SIFT, PolyPhen-2, CADD, and AlphaMissense. The selection of these four tools was made on the basis of their adequate documentation, which is very important given the intended utilisation of the tools in a clinical context. Moreover, these tools are complementary in nature, due to the differing perspectives they employ in their predictions. AlphaMissense takes into account the 3D structure of a reference sequence, while CADD is driven by assumptions concerning the neutrality of a variant in terms of its evolutionary and natural selection processes (Cheng

Table 1: Overview of the number of variants grouped by their functional impact.

Functional impact	**Number of variants**
No function	189
Decreased function	127
Normal function	96
Increased function	7
Unknown or uncertain function	381
Function not assigned	1100

et al., 2023; Schubach et al., 2024). Furthermore, the prediction process of Polyphen-2 is based on sequence, phylogenetic and structural properties that characterise an amino acid change, whereas SIFT is based on the evolutionary conservation of amino acids within protein families (Adzhubei et al., 2013; Vaser et al., 2016).

The final data set used for this evaluation contains the 419 variants that belong to the four relevant classes. Variant effect predictions were performed using Ensembl VEP (version 113), deployed via a Docker (Podman 4.9.4-rhel) container and Apptainer (Version 1.3.2-1). The analysis was aligned to the GRCh38 genome assembly, matching the reference version used in the Ensembl VEP cache. PolyPhen-2 (version 2.2.3) and SIFT (version 6.2.1) prediction scores are incorporated into the Ensembl VEP's output by default. Plugins were used to integrate the additional predictors - CADD (version v1.7) and AlphaMissense (VEP release 113).

Table 2 summarises the score ranges, thresholds and their interpretation. AlphaMissense, CADD and PolyPhen-2 use scales where a higher score indicates that the variant is more likely to be deleterious, damaging or pathogenic. For SIFT, a variant with a score lower than 0.05 is assumed to be deleterious. It should be noted that in the context of CADD, the raw scores are used.

Prediction scores were extracted from the VEP-generated VCF output files using custom scripts. Boxplots and density plots were then created to compare the score distributions across the PharmVar functional impact assigned to the variants. Correlation analysis between the scores of each tool was performed using the Spearman rank correlation metric, because of its robustness to outliers and its non-parametric nature (Rebekić et al., 2015).

The next comparison is based on receiver operating characteristic (ROC) curves and the area under the curve (AUC). Three binary classification scenarios were constructed for each prediction tool to enable variants with no functional impact, decreased functional impact, or increased functional impact to be compared with variants that belong to the class of variants with normal functionality. The tools were evaluated based on the ROC

Table 2: Overview of the variant effect prediction tools by score range, threshold and interpretation.

Tool	Score Range	Threshold	Interpretation
AlphaMissense	[0, 1]	≥ 0.564	Likely pathogenic
		[0.34, 0.564]	Ambiguous
		≤ 0.34	Likely benign
CADD	$\approx$ [–6, 35]	> 10	No strict cutoff, but scores > 10 are often flagged as potentially damaging
PolyPhen-2	[0, 1]	≥ 0.85	Probably damaging
		[0.15, 0.85]	Possibly damaging
		< 0.15	Benign
SIFT	[0, 1]	≤ 0.05	Deleterious
		> 0.05	Tolerated

curves and the AUC, in terms of their ability to separate normal variants from the other functional classes using their prediction scores.

4.3 Ensemble method

A predictive model was developed to investigate whether prediction scores can be used to predict the functional classification of pharmacogenetic variants. To achieve this, the four pathogenicity scoring systems - AlphaMissense, CADD, PolyPhen-2 and SIFT - were used as input features.

As discussed earlier, only seven variants have been assigned to an increased function. This underrepresentation can be attributed to the fact that the occurrence of increased functional impact is most often associated with structural variants, especially duplicates. In order to ensure more balanced classes during the training process of the ensemble method, these variants were excluded. Additionally, variants for which one of the tools could not compute a score were excluded. The final data set comprised 309 variants, all of which constituted complete cases. Table 3 provides an overview of the number of variants assigned to each PharmVar function. These three functional classes served as the target labels for supervised classification.

All scores were processed to align their scales in the same direction, so that higher scores consistently reflect greater predicted deleteriousness. To achieve this, SIFT scores were inverted (i.e. 1 – SIFT) because lower SIFT values indicate a higher probability of a deleterious impact. As discussed in section 4.2, the score range of the predictors varies. The implementation of Z-score normalisation was used to ensure that all features contribute

Table 3: Overview of the number of variants assigned to the selected PharmVar functions.

Function	Number of variants
No function	118
Decreased function	108
Normal function	83

equally to the model. Let s_{ij} be the score of tool i and variant j, then the normalised score $\tilde{s}_{ij}$ is given by

$$\tilde{s}_{ij} = \frac{s_{ij} - \overline{s}_i}{\overline{\sigma}_i}$$

with $i = \{\text{AlphaMissense, CADD, PolyPhen-2}\}$, $\overline{s}_i$ the mean score of tool i and $\overline{\sigma}_i$ the standard deviation of the scores of tool i. For SIFT, the scores first need to be inverted, so the following formula holds

$$\tilde{s}_j = \frac{(1 - s_j) - \overline{s}}{\overline{\sigma}}$$

with $\overline{s}_i$ the mean score of SIFT and $\overline{\sigma}_i$ the standard deviation of the SIFT scores.

The data was divided into a training set and a test set by means of stratified random sampling. This was done to ensure that the functional classes in the training and test sets remained proportional to their distribution in the entire data set. In each functional class, 75% of the variants will be used for the training of the model, with the remaining 25% being retained for the testing phase. This ratio of training data to test data is common practice in the field of machine learning. Table 4 provides an overview of the number of variants assigned to each functional class in the training set and the test set.

Table 4: Overview of the number of variants assigned to each functional class in the training set and test set of the ensemble method.

	No function	Decreased function	Normal function
Training set	88	81	61
Test set	30	27	21

A machine learning model was trained using the scores from the four tools in the training set as input features. First, a multinomial logistic regression model was employed. This model is characterised by its simplicity and the assumption of a linear relationship between the input features and the log-odds of the functional classes. However, the linear relationship is likely to be inadequate for this biologically complex problem. Subsequent

to this, two additional machine learning models were proposed: random forest and support vector machines. Random forest was chosen for its non-parametric nature, it does not assume anything about the correlation between the functional classes and the input features. Support vector machines were trained using a radial kernel, as this effectively handles non-linear relationships between the input features if present. Hyperparameter optimisation was performed for the models based on random forest and support vector machines using 5-fold cross-validation. Once the optimal parameters have been found, the model was trained using the entire training set. Model performance was evaluated on the independent test set using various metrics appropriate for multiclass classification. These included overall accuracy, macro-averaged precision, macro-averaged recall, the F1 score (calculated by averaging the class-specific F1 scores) and ROC curves.

4.4 Computational environment

All variant effect predictions and the Nextflow pipeline generated to summarise the results were executed on the High-Performance Computing (HPC) infrastructure at the KU Leuven/UHasselt Tier-2 clusters. This environment enabled the efficient processing of large variant datasets. To ensure reproducibility across nodes, analyses were performed using Apptainer containers. The output files were then uploaded to R to provide illustrative summaries of the results. The code is available as a GitHub repository: https://github.com/lorepellens/VEP_in_PGx.

5 Results

5.1 Variant effect predictors

The performance of four variant effect predictors (AlphaMissense, CADD, PolyPhen-2 and SIFT) on pharmacovariants was investigated. The results obtained from these predictor tools will be summarised in this section.

The distribution of scores across the PharmVar functional classes was examined to identify any clear differences between the classes for each predictor tool. Figure 8 presents the score distributions using density plots. AlphaMissense exhibited distributions that were all similar in shape and skewed to the left across the 'no function', 'decreased function', and 'normal function' variant classes. The overlap between these distributions indicates limited discriminative power. The curve associated with an increased function displays a different pattern. However, this interpretation should be treated with caution due to the small sample size (n=7) in this class of variants as outlined in section 4.1. A similar pattern was observed for PolyPhen-2 and SIFT. Conversely, CADD score distributions show no clear pattern. One potential explanation for this could be that, in contrast to the scores of the other predictors, the (raw) CADD scores are not scaled.

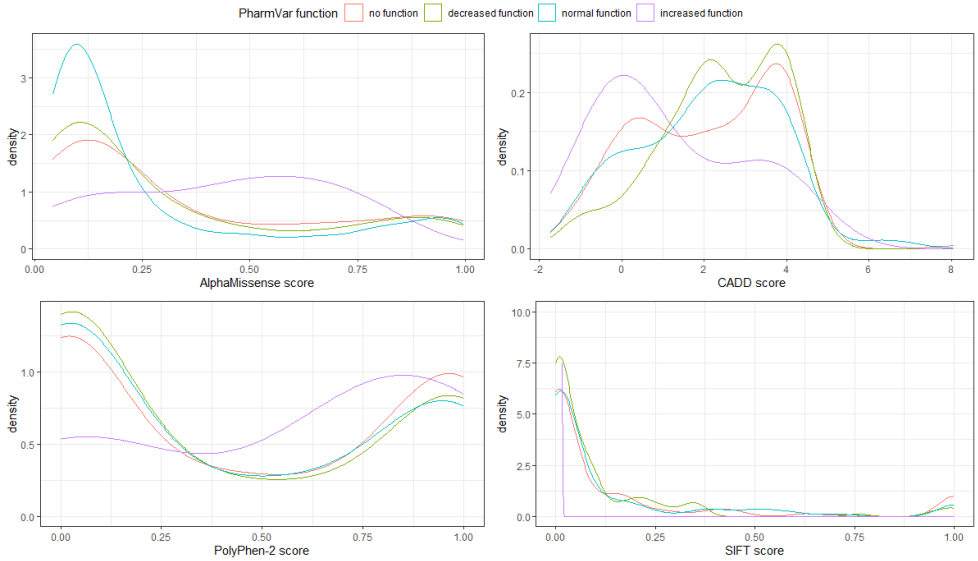

Figure 8: Density plots showing the distribution of scores for variants belonging to the same functional class, for each tool.

An alternative approach to visualise the distribution of prediction scores across the functional classes is through the use of boxplots, as illustrated in Figure 9. Notably, the median scores for the 'no function', 'decreased function' and 'normal function' classes are approximately equal for every tool except for PolyPhen-2. For this tool, while the median scores for 'decreased function' and 'normal function' are still approximately equal to each other, the median score for the 'no function' class is elevated by approximately 0.25. The observed discrepancies between these three classes and the 'increased function' class across all tools may be attributable to the limited sample size of variants with increased functional impact, as discussed previously. The exact median scores and the standard deviation of the functional classes are presented in Table 5 for each tool.

Table 5: The median score (with standard deviation) for the no functional, decreased functional and normal functional classes for each variant effect predictor.

Variant effect predictor	No function	Decreased function	Normal function	Increased function
AlphaMissense	0.197 (0.321)	0.172 (0.309)	0.129 (0.293)	0.462 (0.288)
CADD	2.279 (1.747)	2.449 (1.456)	2.194 (1.674)	-0.105 (1.814)
PolyPhen-2	0.354 (0.441)	0.106 (0.431)	0.089 (0.430)	0.741 (0.478)
SIFT	0.015 (0.307)	0.030 (0.212)	0.020 (0.268)	0.00 (0.006)

A comparative analysis of the tools can be conducted by examining their ROC curves and AUC values. Figure 10 depicts the ROC curves for distinguishing between "normal

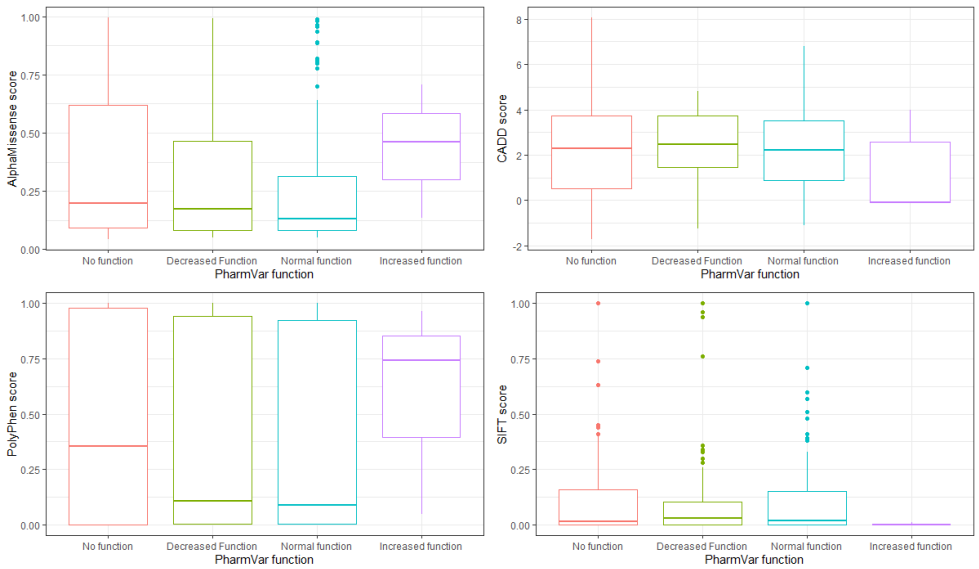

Figure 9: Boxplots showing the distribution of scores for variants belonging to the same functional class, for each tool.

function" and "no function" variants. The results indicate poor classification performance across all tools. The AUC values of AlphaMissense, CADD, PolyPhen-2 and SIFT are equal to 0.58, 0.51, 0.53 and 0.51 respectively. Given that all of these values are close to 0.5, the predictive performance of the tools is barely better than random. AlphaMissense demonstrates superior performance in comparison to the other tools, although this remains poor; its curve shows a slight elevation above the diagonal.

The same approach can be used to evaluate the performance of the tools in distinguishing between "normal function" and "decreased function" variants. This is shown in Figure 11. The AUC values are very similar to those for distinguishing between "normal function" and "no function" variants. They equal 0.53, 0.56, 0.51 and 0.51 for AlphaMissense, CADD, PolyPhen-2 and SIFT respectively. CADD performs best in this case, although the improvement over random classification is only minor.

As demonstrated in Figure 12, the ROC curves show the capacity to distinguish between 'normal function' and 'increased function' variants. The differences between these curves and those presented in the preceding figures, i.e. Figures 10 and 11, were found to be remarkable. This is likely attributable to the fact that these plots are based on only seven variants, thus making it unfeasible to draw any conclusions from them. Further research is required in order to provide detailed information on this functional class. This will necessitate the collection of additional data.

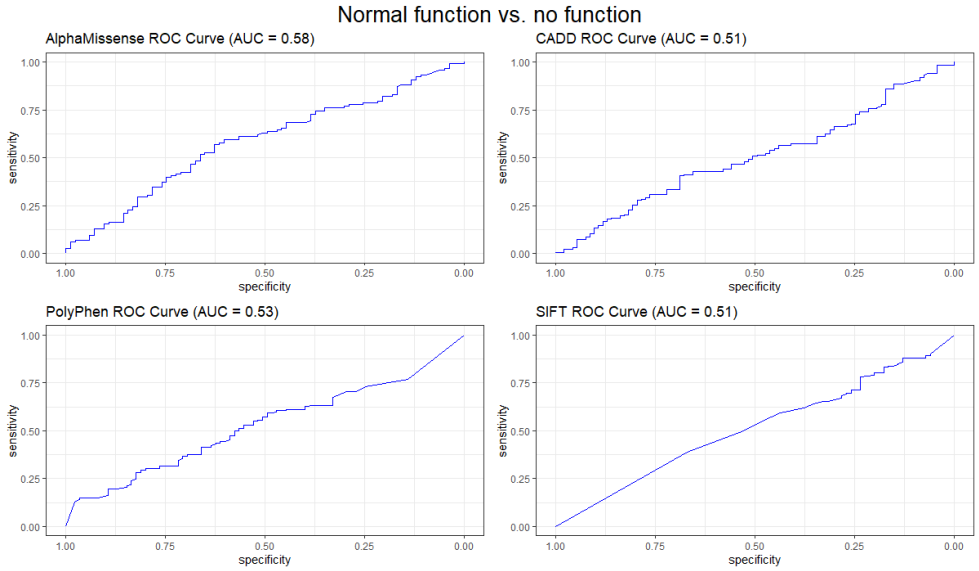

Figure 10: ROC curves showing the ability to distinguish between 'normal function' and 'no function' variants, and AUC for each tool.

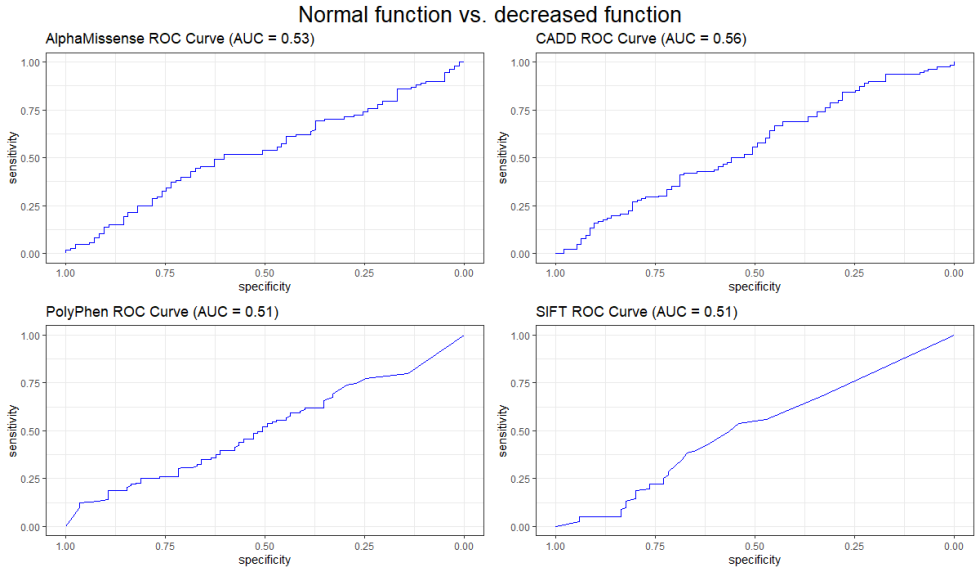

Figure 11: ROC curves showing the ability to distinguish between "normal function" and "decreased function" variants, and AUC for each tool.

Finally, the relationships between the scores from the four tools were examined using Spearman rank correlation analysis. Figure 13 shows the distribution of scores for each tool on the diagonal. Off-diagonal, the correlations between each pair of tools are plotted

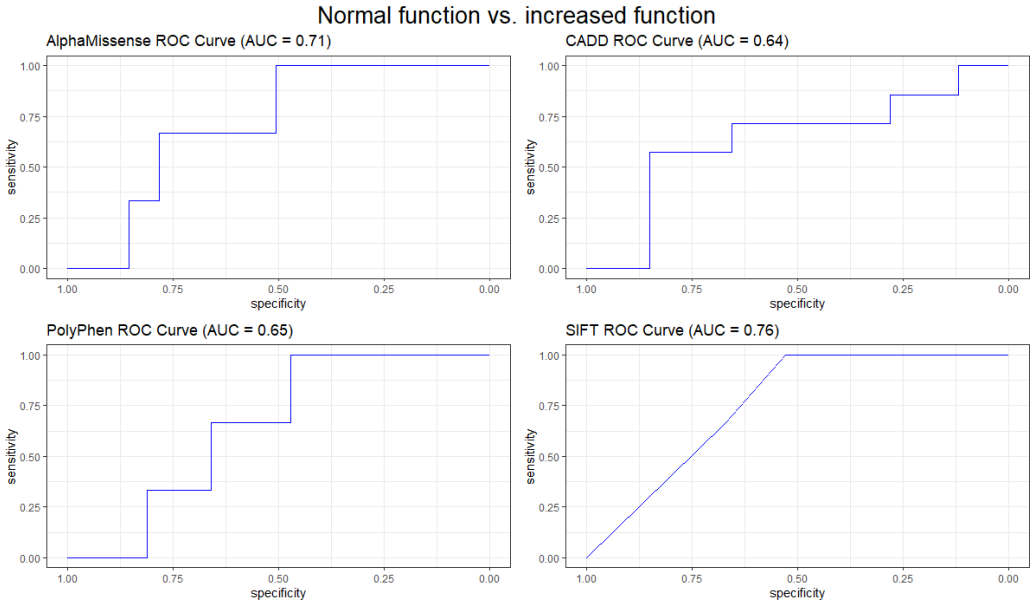

Figure 12: ROC curves showing the ability to distinguish between "normal function" and "increased function" variants, and AUC for each tool.

and summarised in the form of the Spearman correlation coefficient. As outlined in Table 6, the correlations are mainly strong and positive, except for SIFT, which has a strong negative correlation with all the other tools. SIFT scores are interpreted oppositely to those of the other tools which is consistent with its inverse scoring system where lower values indicate higher probability of deleteriousness as seen in Table 2. These strong correlations suggest that multiple tools may provide redundant rather than complementary information. This also explains why their ROC performances were similarly poor: they likely have similar limitations when it comes to distinguishing functional categories in pharmacogenomic contexts. The findings described in this section underscore the need for PGx-specific prediction models, as these tools exhibit limited discriminative potential in differentiating functionally important classes of pharmacovariants.

Table 6: Spearman's rank correlation coefficients for every pair of prediction tools.

	Correlation coefficient
AlphaMissense - CADD	0.737
AlphaMissense - PolyPhen-2	0.736
AlphaMissense - SIFT	-0.796
CADD - PolyPhen-2	0.806
CADD - SIFT	-0.813
PolyPhen-2 - SIFT	-0.799

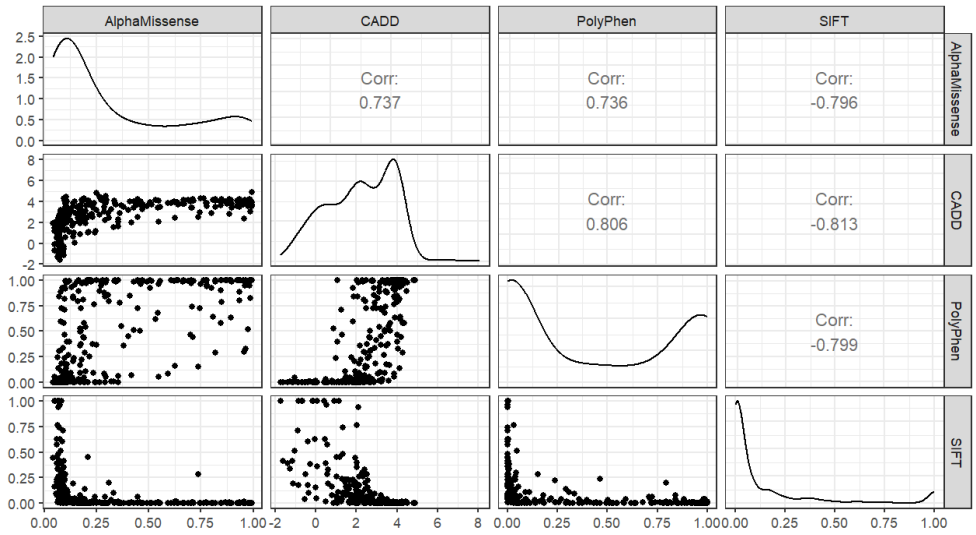

Figure 13: Comparative analysis of the four variant effect predictors AlphaMissense, CADD, PolyPhen-2 and SIFT using Spearman correlations. Negative correlations indicate inverse relationships (e.g. lower SIFT scores correlate with higher AlphaMissense scores)

5.2 Ensemble method

Three ensemble methods were developed as described in section 4.3. The performance of these models are summarised in Table 7.

Table 7: Summary of the performance results of the ensemble method based on multinomial logistic regression, random forest and support vector machine.

Training method	Mean accuracy	Mean recall	Mean precision	Mean F1-score
Multinomial Logistic Regression	34.62%	35.83%	33.02%	32.86%
Random Forest	33.33%	32.43%	31.53%	30.54%
Support Vector Machine	37.18%	37.78%	35.59%	35.33%

Among the three ensemble methods, the multinomial logistic regression model was first evaluated. The model achieved 34.62% accuracy, 35.83% mean recall, 33.02% mean precision and 32.86% mean F1-score. The corresponding ROC curves in Figure 14 demonstrate limited discriminative ability across all functional classes, indicating minimal improvement over random classification.

The ensemble method, which was trained using a random forest, exhibited the poorest performance. The tuning of the parameters was achieved through the implementation of

five-fold cross-validation. The optimal configuration consisted of one randomly selected feature (score) per split, a nodesize equal to 9 and 750 decision trees. The model demonstrated a 33.33% level of accuracy accompanied by mean precision, mean recall and mean F1-score equal to 32.43%, 31.53% and 30.54% respectively. These metrics indicate that the model trained with random forest performs only marginally worse than the multinomial logistic regression model.

Support Vector Machine demonstrates the most effective categorisation across all functional categories. The cost and gamma value were chosen to be equal to 32 and 0.0625 respectively based on the results of five-fold cross-validation. As demonstrated in Figure 14, the ROC curves obtained from this model are, for the majority of values of sensitivity and specificity, situated well above the diagonal line of random classification. This method achieves the highest evaluation metrics: 37.18% accuracy, 37.78% mean recall, 35.59% mean precision, and 35.33% mean F1 score.

Among the ensemble methods evaluated, the support vector machines demonstrated the best performance in classifying pharmacovariants by functional impact. It outperformed both multinomial logistic regression and random forest, although the overall predictive accuracy remained modest.

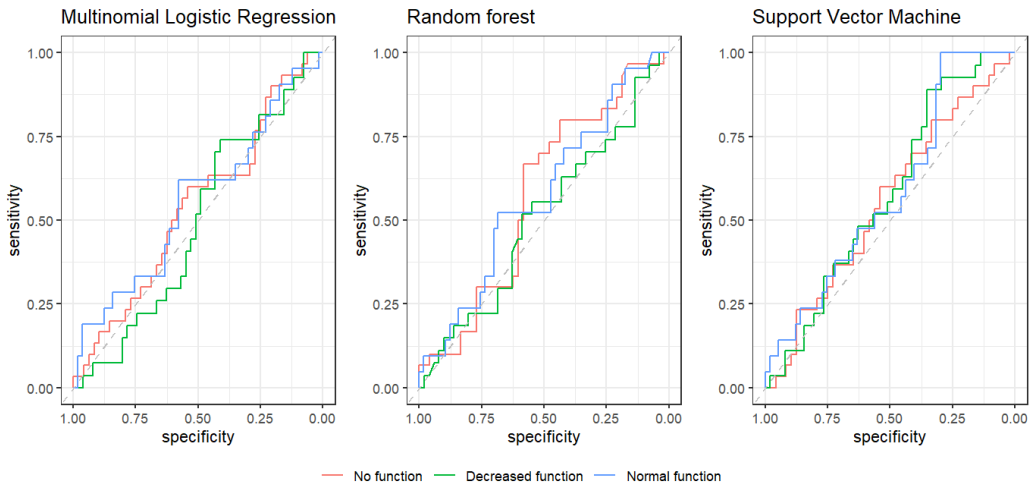

Figure 14: ROC curves for the three ensemble methods colored by functional class (left: multinomial logistic regression, middle: random forest, right: support vector machines).

5.3 Nextflow pipeline

A Nextflow (version 23.10.0) pipeline was developed to systematically annotate genetic variants and assess their potential functional impact. This pipeline combines multiple

variant effect predictors with the ensemble classification model based on support vector machines developed in this thesis. The pipeline takes a VCF file as input for the variant effect predictors. Figure 15 illustrates a schematic overview of the pipeline, with each module described below.

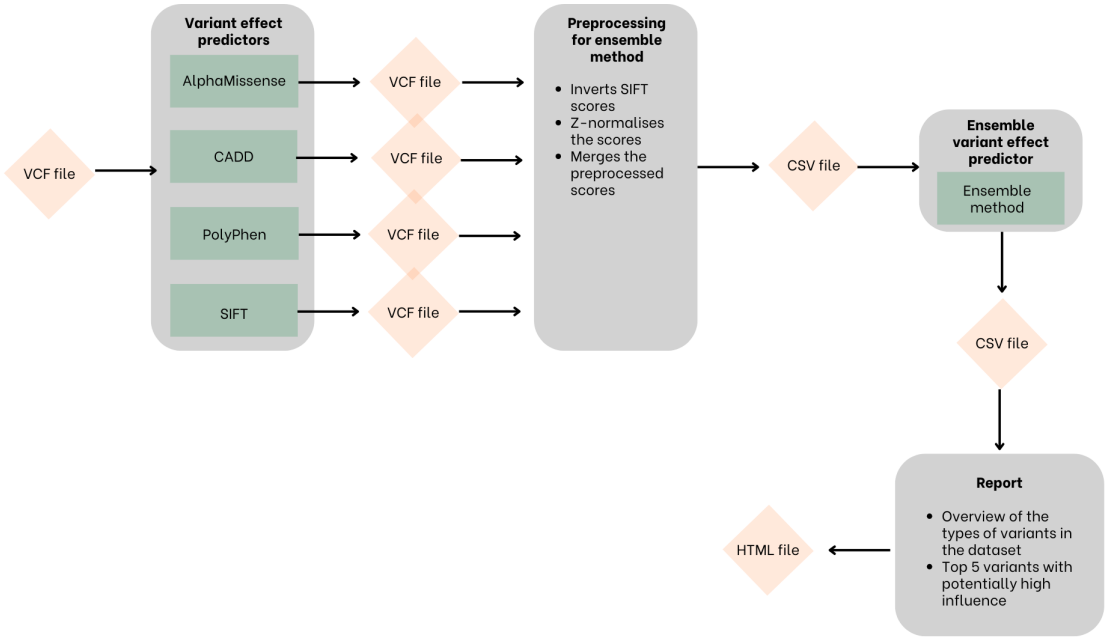

Figure 15: Diagram of the nextflow pipeline.

Variant effect predictors
Each variant is annotated using the four variant effect predictors described in section 4.2. The scores generated by each tool are stored in a VCF file with extra information on each variant such as the chromosome number and rs identifier.

Preprocessing for ensemble method
The four VCF files obtained from the variant effect predictors are preprocessed by transforming the scores into the correct format to be used as input of the ensemble method. The process involves first inverting the SIFT scores, and subsequently normalising all scores using Z-normalisation. Following this, the scores are aggregated into a single CSV file.

Ensemble variant effect predictor
The CSV file containing the scores of each variant will be used to predict the functional impact of the variants present in the file directly, using the Ensemble method based on support vector machines developed as described in section 5.2. The output is a CSV file containing the rs identifier and PharmVar function for each variant.

Report generation
Finally, the pipeline compiles the results into a structured HTML report. This includes an overview of the types of variants in the first VCF file and a prioritised list of the top five potentially high-impact variants. This automated pipeline ensures a consistent, reproducible and interpretable assessment of variant function, focusing on PGx.

6 Discussion

This thesis evaluated the applicability of four widely used variant effect predictors - AlphaMissense, CADD, PolyPhen-2 and SIFT - in the context of PGx. While the literature supports that these tools can identify disease-associated variants robustly, they are limited in their ability to predict the functional impact of pharmacovariants. This highlights a fundamental challenge: PGx requires predictions that align with drug metabolism phenotypes rather than pathogenicity scores. Consequently, models trained for pathogenicity prediction tend to underperform when applied to PGx. This finding aligns with recent literature highlighting the need for PGx-specific tools.

The density plots and ROC curves in Section 5.1 further confirmed these shortcomings. The evaluated predictors failed to distinguish between the functional classes of pharmacovariants, as defined by PharmVar. The resulting AUC scores had a range between 0.51 and 0.58 across all tools, which is barely better than random classification. These findings reinforce the idea that PGx is a domain-specific challenge that these tools were not designed to address. Consequently, the conversion of tool scores into metaboliser classes is a challenging task. Moreover, the strong correlations observed between the tools indicate their capacity to detect the same underlying biological signal in the data: deleteriousness. This may contribute to their limited predictive ability, as they may be trained to make similar, non-PGx-specific decisions when classifying variants.

In order to ascertain whether an ensemble method could lead to improvement, it was developed by combining scores from the individual tools in a multiclass classification framework. Among the tested models, the support vector machine demonstrated superior performance in comparison to the multinomial logistic regression and random forest models. The support vector machine attained an accuracy rate of 37.18%, a precision of 35.59% and a recall of 37.78%. The ensemble model that was trained using random forest demonstrated the poorest performance. One potential explanation for this could be that random forest tends to underperform if the feature space is not diverse or contains redundant signals. Multinomial logistic regression is also sensitive to multicollinearity, which can inflate variance and reduce performance.

The moderate performance of the support vector machine ensemble method highlights the difficulty of PGx classification using general-purpose predictors. Despite the fact that the

overall predictive accuracy remained modest, it is hypothesised that performance gains can be achieved by integrating biological domain knowledge into the models. For instance, variant effect predictors like APF2 (Activity Prediction Framework 2) have demonstrated success by integrating gene-specific attributes, drug-pathway data. APF2, developed by Zhou et al. (2024), was trained to predict enzyme activity using curated activity scores, achieving a reported performance of over 80% in distinguishing functional classes in CYP2D6. While this thesis relied solely on general-purpose predictors, it is promising that a relatively simple ensemble method could yield clinically relevant insights. Nonetheless, leveraging domain-specific features may be essential for achieving robust PGx predictions. A potential constraint to consider is that the data sets used for training and evaluating the ensemble method were limited due to missing scores or underrepresented functional classes. This may have affected the generalisability of the model. The issue of restricted access to data resources represents a common challenge in the context of PGx studies.

The development of the Nextflow pipeline adds practical value by turning the findings into a reproducible, scalable workflow. It integrates the four evaluated predictors and the ensemble method, and supports VCF input. Although this level of automation is still in its early stages of development, it is nevertheless a key step towards routine PGx interpretation and clinical decision support. The pipeline delineated in this thesis will be expanded upon subsequent analysis, incorporating recommendations concerning pharmacogenomic relevance and implications for drug dosage. This expanded pipeline will then be integrated in a clinical workflow.

7 Ethical thinking

As PGx research advances and becomes integrated into clinical practice, ethical challenges arise that must be carefully considered. These challenges relate to data privacy, genetic discrimination, incidental findings and informed consent.

This thesis used publicly available variant data obtained from PharmVar, and did not involve patient-specific or identifiable genomic data. Nevertheless, any research involving genetic information must acknowledge the sensitivity of such data. PGx testing involves sensitive genetic information and can reveal potential responses to drugs, predispositions to diseases, and inherited traits. If the predictive frameworks evaluated here were to be applied in clinical settings, robust data protection standards would be required to prevent misuse, discrimination or breaches of privacy.

Whole-genome and exome sequencing are often used in PGx and may reveal incidental or secondary findings that are unrelated to the drug response being studied. These could include an increased risk of developing untreatable diseases or the discovery of non-paternity. Ethical frameworks must therefore guide decisions on what should be reported

back to patients.

Given the hereditary nature of genetic variants, pharmacogenomic testing often provides insights not only about an individual, but also about their biological relatives. This raises complex ethical questions regarding the obligation to inform family members of results that could affect their health. For instance, a variant affecting drug metabolism in one person may suggest comparable risks for their siblings or offspring. However, the patient's right to confidentiality must be balanced against the potential benefit to relatives. Clear guidelines are needed to navigate these issues, particularly in cases where the patient does not wish to share information.

A lack of transparency in variant effect predictors present additional ethical concerns. There is a clear need for improvement in this ethical consideration. In many cases, the pipelines and data sets used by the predictors are not open-source. This hinders the evaluation of the models and comparison with new tools.

In order to be implemented in a clinical setting, variant effect predictors must demonstrate a high degree of accuracy. Despite this, these predictors are still at risk of misclassifying a variant. This underscores the necessity for clear guidelines concerning the clinical accountability of healthcare providers in cases where patients are not administered the most efficient pharmaceutical agent as a result of the use of prediction tools.

The variants analysed in this thesis were sourced from databases which may reflect population biases, since certain populations (e.g., people of African, Indigenous, or mixed ancestry) may be underrepresented in genomic studies. Consequently, the findings and predictive models developed here may not be applicable to all ethnic groups. This raises a critical ethical concern regarding the potential for unequal benefit, or even harm, if tools trained on non-representative data are deployed in clinical settings. It is therefore essential to ensure diverse datasets and equitable access to PGx testing in order to avoid biased clinical decisions.

8 Societal relevance and stakeholders

PGx has broad societal impact due to its intersection of genomics, healthcare, and bioinformatics. As a result, PGx is highly relevant to a variety of sectors and communities. As the cost of genomic sequencing continues to decrease, individualised drug therapy can quickly become a clinical reality.

The enhanced prediction of novel or rare variants has the power to reduce ADRs. This has the potential to reduce both hospital admissions and medication changes, as well as a decrease the need for emergency room visits. Overall, the incorporation of novel variants

could result in lower healthcare costs through improved treatment outcomes.

The implementation of pharmacogenomic frameworks holds the promise of enhancing healthcare services within hospital settings. The long-term goal of the Jessa Hospital is to offer pre-emptive care through WGS. This thesis contributes to the growing understanding of PGx as an important application of WGS, with the potential to advance the implementation of pre-emptive care in clinical settings, particularly for patients with complex pharmacogenomic profiles involving novel or rare variants.

Provided they are implemented carefully and responsibly, PGx-specific variant effect predictors have the potential to guide healthcare professionals and patients towards more informed treatment decisions. Patients stand to benefit from the improved drug safety and efficacy. Given that healthcare professionals such as doctors, will use the predictors for decision-making purposes, it is important that they have sufficient knowledge about them. This necessitates investment in pharmacogenomic education that targets both healthcare providers and patients alike.

The ongoing development and evaluation of PGx models will be of interest to researchers and bioinformaticians, who play a critical role in advancing the field. At the same time, it is crucial for regulatory and ethical bodies to ensure that predictive models are transparent, fair and validated across diverse populations. Additionally, industry stakeholders, including pharmaceutical companies and genomic service providers, are increasingly incorporating PGx into drug discovery and diagnostics. By bridging technical performance with clinical relevance, the interdisciplinary nature of PGx research has the potential to shape the future of personalised medicine.

9 Concluding remarks and future research

This thesis explored the landscape of variant effect prediction within PGx, emphasising the importance of accurately assessing genetic variants for personalised medicine. By evaluating different prediction tools and creating ensemble methods, it became clear that combining multiple predictors can improve the reliability of predictions, thus facilitating more informed clinical decision-making.

The ensemble method, which was constructed based on the scores of AlphaMissense, CADD, PolyPhen-2 and SIFT, showed that training a combination of disease-related variant predictors along with PharmVar functional annotations has moderate performance in predicting the function of pharmacovariants. This suggests that using more domain-specific tools, such as APF2, could yield improved results; however, further validation across diverse populations remains essential.

Future research should focus on several key areas. Firstly, integrating functional genomic

data such as transcriptomics and proteomics could improve our understanding of variant impacts. Secondly, developing machine learning models that are trained using large, well-annotated pharmacogenomic datasets could enhance the precision and broad applicability of the predictor. Thirdly, standardising variant interpretation guidelines will be crucial in facilitating clinical adoption. Finally, pharmacogenomic education for healthcare professionals will be necessary to ensure safe use of variant effect predictors.

In conclusion, the ongoing evolution of variant effect prediction methodologies offers great potential for the future of personalised pharmacotherapy. By encouraging interdisciplinary collaboration and making use of emerging technologies, the PGx community can bring the full potential of precision medicine closer to realisation, ultimately improving therapeutic outcomes and minimising ADRs.

References

Ghloamareza Abdi, Maryam Abbasi Tarighat, Mukul Jain, Reshma Tendulkar, Mugdha Tendulkar, and Mukul Barwant. Revolutionizing Genomics: Exploring the Potential of Next-Generation Sequencing. In Vijai Singh and Ajay Kumar, editors, *Advances in Bioinformatics*, pages 1–33. Springer Nature, Singapore, 2024. ISBN 978-981-99-8401-5. doi: 10.1007/978-981-99-8401-5_1. URL https://doi.org/10.1007/978-981-99-8401-5_1.

Ivan Adzhubei, Daniel M. Jordan, and Shamil R. Sunyaev. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]*, 0 7:Unit7.20, January 2013. ISSN 1934-8266. doi: 10.1002/0471142905.hg0720s76. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4480630/.

Maryam Jamal AlSaeed, Peter Ramdhan, Jean Gabriel Malave, Islam Eljilany, Taimour Langaee, Caitrin W. McDonough, Gustavo Seabra, Chenglong Li, and Larisa H. Cavallari. Assessing the Performance of In silico Tools and Molecular Dynamics Simulations for Predicting Pharmacogenetic Variant Impact. *Clinical Pharmacology & Therapeutics*, 116(4):1082–1089, 2024. ISSN 1532-6535. doi: 10.1002/cpt.3348. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.3348. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpt.3348.

Ursula Amstutz, Linda M. Henricks, Steven M. Offer, Julia Barbarino, Jan H.M. Schellens, Jesse J. Swen, Teri E. Klein, Howard L. McLeod, Kelly E. Caudle, Robert B. Diasio, and Matthias Schwab. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clinical Pharmacology & Therapeutics*, 103(2):210–216, February 2018. ISSN 0009-9236, 1532-6535. doi: 10.1002/cpt.911. URL https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.911.

Oluchukwu O Anunobi. Pharmacogenomics as a Tool in Addressing Genetic Variation-Dependent Adverse Drug Reactions. *Dutse Journal of Pure and Applied Sciences*, 10(1b): 37–54, 2024.

Mariamena Arbitrio, Maria Teresa Di Martino, Francesca Scionti, Vito Barbieri, Licia Pensabene, and Pierosandro Tagliaferri. Pharmacogenomic Profiling of ADME Gene Variants: Current Challenges and Validation Perspectives. *High-Throughput*, 7(4):40, December 2018. ISSN 2571-5135. doi: 10.3390/ht7040040. URL https://www.mdpi.com/2571-5135/7/4/40. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

Christina A. Austin-Tse, Vaidehi Jobanputra, Denise L. Perry, David Bick, Ryan J. Taft, Eric Venner, Richard A. Gibbs, Ted Young, Sarah Barnett, John W. Belmont, Nicole Boczek, Shimul Chowdhury, Katarzyna A. Ellsworth, Saurav Guha, Shashikant Kulkarni, Cherisse Marcou, Linyan Meng, David R. Murdock, Atteeq U. Rehman, Elizabeth Spiteri, Amanda Thomas-Wilson, Hutton M. Kearney, Heidi L. Rehm, and Med-

ical Genome Initiative*. Best practices for the interpretation and reporting of clinical whole genome sequencing. *npj Genomic Medicine*, 7(1):27, April 2022. ISSN 2056-7944. doi: 10.1038/s41525-022-00295-z. URL https://www.nature.com/articles/s41525-022-00295-z.

Chiara Auwerx, Marie C. Sadler, Alexandre Reymond, and Zoltán Kutalik. From pharmacogenetics to pharmaco-omics: Milestones and future directions. *Human Genetics and Genomics Advances*, 3(2):100100, April 2022. ISSN 26662477. doi: 10.1016/j.xhgg.2022.100100. URL https://linkinghub.elsevier.com/retrieve/pii/S2666247722000161.

Frederik Otzen Bagger, Line Borgwardt, Andreas Sand Jespersen, Anna Reimer Hansen, Birgitte Bertelsen, Miyako Kodama, and Finn Cilius Nielsen. Whole genome sequencing in clinical practice. *BMC Medical Genomics*, 17(1):39, January 2024. ISSN 1755-8794. doi: 10.1186/s12920-024-01795-w. URL https://doi.org/10.1186/s12920-024-01795-w.

Julia M. Barbarino, Christine E. Staatz, Raman Venkataramanan, Teri E. Klein, and Russ B. Altman. PharmGKB summary: cyclosporine and tacrolimus pathways. *Pharmacogenetics and genomics*, 23(10):563–585, October 2013. ISSN 1744-6872. doi: 10.1097/FPC.0b013e328364db84. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4119065/.

Saurabh Belsare, Michal Levy-Sakin, Yulia Mostovoy, Steffen Durinck, Subhra Chaudhuri, Ming Xiao, Andrew S Peterson, Pui-Yan Kwok, Somasekar Seshagiri, and Jeffrey D Wall. Evaluating the quality of the 1000 genomes project data. *BMC genomics*, 20:1–14, 2019. Publisher: Springer.

Julie Berg, Gathi Abraham, Adelaide Robb, and Finza Latif. Chapter 30 - Pediatric psychiatric disorders. In Dennis Dietzen, Michael Bennett, Edward Wong, and Shannon Haymond, editors, *Biochemical and Molecular Basis of Pediatric Disease (Fifth Edition)*, pages 1057–1092. Academic Press, January 2021. ISBN 978-0-12-817962-8. doi: 10.1016/B978-0-12-817962-8.00033-0. URL https://www.sciencedirect.com/science/article/pii/B9780128179628000330.

Ka Birdwell, B Decker, Jm Barbarino, Jf Peterson, Cm Stein, W Sadee, D Wang, Aa Vinks, Y He, Jj Swen, Js Leeder, Rhn Van Schaik, Ke Thummel, Te Klein, Ke Caudle, and Iam MacPhee. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for *CYP3A5* Genotype and Tacrolimus Dosing. *Clinical Pharmacology & Therapeutics*, 98 (1):19–24, July 2015. ISSN 0009-9236, 1532-6535. doi: 10.1002/cpt.113. URL https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.113. Publisher: Wiley.

Ewan Birney and Nicole Soranzo. The end of the start for population sequencing. *Nature*, 526(7571):52–53, October 2015. ISSN 1476-4687. doi: 10.1038/526052a. URL https://www.nature.com/articles/526052a. Publisher: Nature Publishing Group.

Chad A. Bousman, James M. Stevenson, Laura B. Ramsey, Katrin Sangkuhl, J. Kevin Hicks, Jeffrey R. Strawn, Ajeet B. Singh, Gualberto Ruaño, Daniel J. Mueller, Evangelia Eirini

Tsermpini, Jacob T. Brown, Gillian C. Bell, J. Steven Leeder, Andrea Gaedigk, Stuart A. Scott, Teri E. Klein, Kelly E. Caudle, and Jeffrey R. Bishop. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for *CYP2D6* , *CYP2C19* , *CYP2B6* , *SLC6A4* , and *HTR2A* Genotypes and Serotonin Reuptake Inhibitor Antidepressants. *Clinical Pharmacology & Therapeutics*, 114(1):51–68, July 2023. ISSN 0009-9236, 1532-6535. doi: 10.1002/cpt.2903. URL https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.2903.

Sharon R. Browning and Brian L. Browning. Haplotype phasing: existing methods and new developments. *Nature Reviews Genetics*, 12(10):703–714, October 2011. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg3054. URL https://www.nature.com/articles/nrg3054.

Jeffrey S. Buguliskis. Pharmacogenomics *Serves as the Critical Driver for Precision Medicine*. *Clinical OMICs*, 2(6):12–14, 16, June 2015. ISSN 2334-1351. doi: 10.1089/clinomi.02.06.06. URL http://www.liebertpub.com/doi/10.1089/clinomi.02.06.06.

Kendall J. Burdick, Joy D. Cogan, Lynette C. Rives, Amy K. Robertson, Mary E. Koziura, Elly Brokamp, Laura Duncan, Vickie Hannig, Jean Pfotenhauer, Rena Vanzo, Michael S. Paul, Anna Bican, Thomas Morgan, Jessica Duis, John H. Newman, Rizwan Hamid, and John A. Phillips. Limitations of exome sequencing in detecting rare and undiagnosed diseases. *American journal of medical genetics. Part A*, 182(6):1400–1406, June 2020. ISSN 1552-4825. doi: 10.1002/ajmg.a.61558. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8057342/.

Weimin Cai, Zhaoqian Liu, Liyan Miao, and Xiaoqiang Xiang, editors. *Pharmacogenomics in Precision Medicine: From a Perspective of Ethnic Differences*. Springer Singapore, Singapore, 2020. ISBN 978-981-15-3894-0 978-981-15-3895-7. doi: 10.1007/978-981-15-3895-7. URL http://link.springer.com/10.1007/978-981-15-3895-7.

Pedro J Caraballo, Joseph A Sutton, Jyothsna Giri, Jessica A Wright, Wayne T Nicholson, Iftikhar J Kullo, Mark A Parkulo, Suzette J Bielinski, and Ann M Moyer. Integrating pharmacogenomics into the electronic health record by implementing genomic indicators. *Journal of the American Medical Informatics Association : JAMIA*, 27 (1):154–158, October 2019. ISSN 1067-5027. doi: 10.1093/jamia/ocz177. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913212/.

PJ Caraballo, SJ Bielinski, JL St. Sauver, and RM Weinshilboum. Electronic medical record-integrated pharmacogenomics and related clinical decision support concepts. *Clinical Pharmacology & Therapeutics*, 102(2):254–264, 2017. Publisher: Wiley Online Library.

Yoseph Caraco. Genes and the Response to Drugs. *New England Journal of Medicine*, 351(27): 2867–2869, December 2004. ISSN 0028-4793. doi: 10.1056/NEJMe048278. URL https://www.nejm.org/doi/full/10.1056/NEJMe048278. Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMe048278.

Daniel Carranza-Leon, Alyson L. Dickson, Andrea Gaedigk, C. Michael Stein, and Cecilia P. Chung. CYP2D6 genotype and reduced codeine analgesic effect in real-world clinical practice. *The Pharmacogenomics Journal*, 21(4):484–490, August 2021. ISSN 1473-1150. doi: 10.1038/s41397-021-00226-8. URL https://www.nature.com/articles/s41397-021-00226-8. Publisher: Nature Publishing Group.

Kelly E. Caudle, Katrin Sangkuhl, Michelle Whirl-Carrillo, Jesse J. Swen, Cyrine E. Haidar, Teri E. Klein, Roseann S. Gammal, Mary V. Relling, Stuart A. Scott, Daniel L. Hertz, Henk-Jan Guchelaar, and Andrea Gaedigk. Standardizing 2D6 Genotype to Phenotype Translation: Consensus Recommendations from the Clinical Pharmacogenetics Implementation Consortium and Dutch Pharmacogenetics Working Group. *Clinical and Translational Science*, 13(1):116–124, 2020. ISSN 1752-8062. doi: 10.1111/cts.12692. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cts.12692. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cts.12692.

Rosane Charlab and Lei Zhang. Pharmacogenomics: Historical Perspective and Current Status. In Federico Innocenti and Ron H.N. van Schaik, editors, *Pharmacogenomics: Methods and Protocols*, pages 3–22. Humana Press, Totowa, NJ, 2013. ISBN 978-1-62703-435-7. doi: 10.1007/978-1-62703-435-7_1. URL https://doi.org/10.1007/978-1-62703-435-7_1.

Jun Cheng, Guido Novati, Joshua Pan, Clare Bycroft, Akvilė Žemgulytė, Taylor Applebaum, Alexander Pritzel, Lai Hong Wong, Michal Zielinski, Tobias Sargeant, Rosalia G. Schneider, Andrew W. Senior, John Jumper, Demis Hassabis, Pushmeet Kohli, and Žiga Avsec. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*, 381(6664):eadg7492, September 2023. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adg7492. URL https://www.science.org/doi/10.1126/science.adg7492.

Yongwook Choi, Agnes P. Chan, Ewen Kirkness, Amalio Telenti, and Nicholas J. Schork. Comparison of phasing strategies for whole human genomes. *PLoS Genetics*, 14(4):e1007308, April 2018. ISSN 1553-7390. doi: 10.1371/journal.pgen.1007308. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5903673/.

Nadine Cohen and Y. James Kang, editors. *Pharmacogenomics and Personalized Medicine*. Methods in Pharmacology and Toxicology. Humana Press, Totowa, NJ, 2008. ISBN 978-1-934115-04-6 978-1-59745-439-1. doi: 10.1007/978-1-59745-439-1. URL http://link.springer.com/10.1007/978-1-59745-439-1.

Gregory M. Cooper and Jay Shendure. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Reviews Genetics*, 12(9):628–640, September 2011. ISSN 1471-0064. doi: 10.1038/nrg3046. URL https://www.nature.com/articles/nrg3046. Publisher: Nature Publishing Group.

Rhonda M. Cooper-DeHoff, Mikko Niemi, Laura B. Ramsey, Jasmine A. Luzum, E. Katriina Tarkiainen, Robert J. Straka, Li Gong, Sony Tuteja, Russell A. Wilke, Mia

Wadelius, Eric A. Larson, Dan M. Roden, Teri E. Klein, Sook Wah Yee, Ronald M. Krauss, Richard M. Turner, Latha Palaniappan, Andrea Gaedigk, Kathleen M. Giacomini, Kelly E. Caudle, and Deepak Voora. The Clinical Pharmacogenetics Implementation Consortium Guideline for *SLCO1B1*, *ABCG2*, and *CYP2C9* genotypes and Statin-Associated Musculoskeletal Symptoms. *Clinical Pharmacology & Therapeutics*, 111 (5):1007–1021, May 2022. ISSN 0009-9236, 1532-6535. doi: 10.1002/cpt.2557. URL https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.2557. Publisher: Wiley.

Kristine R. Crews, Andrew A. Monte, Rachel Huddart, Kelly E. Caudle, Evan D. Kharasch, Andrea Gaedigk, Henry M. Dunnenberger, J. Steven Leeder, John T. Callaghan, Caroline Flora Samer, Teri E. Klein, Cyrine E. Haidar, Sara L. Van Driest, Gualberto Ruano, Katrin Sangkuhl, Larisa H. Cavallari, Daniel J. Müller, Cynthia A. Prows, Mohamed Nagy, Andrew A. Somogyi, and Todd C. Skaar. Clinical Pharmacogenetics Implementation Consortium Guideline for *CYP2D6* , *OPRM1* , and *COMT* Genotypes and Select Opioid Therapy. *Clinical Pharmacology & Therapeutics*, 110(4):888–896, October 2021. ISSN 0009-9236, 1532-6535. doi: 10.1002/cpt.2149. URL https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.2149.

Fiona Cunningham, James E Allen, Jamie Allen, Jorge Alvarez-Jarreta, M Ridwan Amode, Irina M Armean, Olanrewaju Austine-Orimoloye, Andrey G Azov, If Barnes, Ruth Bennett, Andrew Berry, Jyothish Bhai, Alexandra Bignell, Konstantinos Billis, Sanjay Boddu, Lucy Brooks, Mehrnaz Charkhchi, Carla Cummins, Luca Da Rin Fioretto, Claire Davidson, Kamalkumar Dodiya, Sarah Donaldson, Bilal El Houdaigui, Tamara El Naboulsi, Reham Fatima, Carlos Garcia Giron, Thiago Genez, Jose Gonzalez Martinez, Cristina Guijarro-Clarke, Arthur Gymer, Matthew Hardy, Zoe Hollis, Thibaut Hourlier, Toby Hunt, Thomas Juettemann, Vinay Kaikala, Mike Kay, Ilias Lavidas, Tuan Le, Diana Lemos, José Carlos Marugán, Shamika Mohanan, Aleena Mushtaq, Marc Naven, Denye N Ogeh, Anne Parker, Andrew Parton, Malcolm Perry, Ivana Piližota, Irina Prosovetskaia, Manoj Pandian Sakthivel, Ahamed Imran Abdul Salam, Bianca M Schmitt, Helen Schuilenburg, Dan Sheppard, José G Pérez-Silva, William Stark, Emily Steed, Kyösti Sutinen, Ranjit Sukumaran, Dulika Sumathipala, Marie-Marthe Suner, Michal Szpak, Anja Thormann, Francesca Floriana Tricomi, David Urbina-Gómez, Andres Veidenberg, Thomas A Walsh, Brandon Walts, Natalie Willhoft, Andrea Winterbottom, Elizabeth Wass, Marc Chakiachvili, Bethany Flint, Adam Frankish, Stefano Giorgetti, Leanne Haggerty, Sarah E Hunt, Garth R IIsley, Jane E Loveland, Fergal J Martin, Benjamin Moore, Jonathan M Mudge, Matthieu Muffato, Emily Perry, Magali Ruffier, John Tate, David Thybert, Stephen J Trevanion, Sarah Dyer, Peter W Harrison, Kevin L Howe, Andrew D Yates, Daniel R Zerbino, and Paul Flicek. Ensembl 2022. *Nucleic Acids Research*, 50(D1):D988–D995, January 2022. ISSN 0305-1048. doi: 10.1093/nar/gkab1049. URL https://doi.org/10.1093/nar/gkab1049.

AK Daly, J Brockmoller, F Broly, M Eichelbaum, WE Evans, FJ Gonzalez, J-D Huang, JR Idle, M Ingelman-Sundberg, T Ishizaki, and others. Nomenclature for human CYP2D6 alleles. *Pharmacogenetics and Genomics*, 6(3):193–201, 1996. Publisher: LWW.

Brittney H. Davis and Nita A. Limdi. Translational Pharmacogenomics: Discovery, Evidence Synthesis and Delivery of Race-Conscious Medicine. *Clinical Pharmacology & Therapeutics*, 110(4):909–925, 2021. ISSN 1532-6535. doi: 10.1002/cpt.2357. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.2357. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpt.2357.

Laura Dean and Megan Kane. Codeine therapy and CYP2D6 genotype. 2021.

Karen Eilbeck, Suzanna E Lewis, Christopher J Mungall, Mark Yandell, Lincoln Stein, Richard Durbin, and Michael Ashburner. The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biology*, 6(5):R44, 2005. ISSN 1474-7596. doi: 10.1186/gb-2005-6-5-r44. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1175956/.

Dietmar Enko, Simon Michaelis, Christopher Schneider, Erich Schaflinger, Andreas Baranyi, Wolfgang J Schnedl, and Daniel J Muller. The Use of Next-Generation Sequencing in Pharmacogenomics. *Clinical laboratory*, 69(8), 2023.

Francisco Esteves, José Rueff, and Michel Kranendonk. The central role of cytochrome P450 in xenobiotic metabolism—a brief review on a fascinating enzyme family. *Journal of xenobiotics*, 11(3):94–114, 2021. Publisher: MDPI.

Danielle V. Eusuf and Elizabeth Thomas. Pharmacokinetic variation. *Anaesthesia & Intensive Care Medicine*, 20(2):126–129, February 2019. ISSN 1472-0299, 1878-7584. doi: 10.1016/j.mpaic.2018.12.006. URL https://www.anaesthesiajournal.co.uk/article/S1472-0299(18)30281-9/fulltext. Publisher: Elsevier.

Brent Ewing and Phil Green. Base-Calling of Automated Sequencer Traces Using *Phred*. II. Error Probabilities. *Genome Research*, 8(3):186–194, March 1998. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.8.3.186. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.8.3.186.

A Gaedigk, Sd Simon, Re Pearce, Ld Bradford, Mj Kennedy, and Js Leeder. The CYP2D6 Activity Score: Translating Genotype Information into a Qualitative Measure of Phenotype. *Clinical Pharmacology & Therapeutics*, 83 (2):234–242, 2008. ISSN 1532-6535. doi: 10.1038/sj.clpt.6100406. URL https://onlinelibrary.wiley.com/doi/abs/10.1038/sj.clpt.6100406. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1038/sj.clpt.6100406.

Andrea Gaedigk, Katrin Sangkuhl, Michelle Whirl-Carrillo, Teri Klein, and J. Steven Leeder. Prediction of CYP2D6 phenotype from genotype across world populations. *Genetics in Medicine*, 19(1):69–76, January 2017. ISSN 1098-3600, 1530-0366. doi: 10.1038/gim.2016.80. URL https://www.gimjournal.org/article/S1098-3600(21)01541-0/fulltext. Publisher: Elsevier.

Andrea Gaedigk, Katrin Sangkuhl, Michelle Whirl-Carrillo, Greyson P. Twist, Teri E. Klein, Neil A. Miller, and the PharmVar Steering Committee. The Evolution of PharmVar. *Clinical Pharmacology & Therapeutics*, 105(1):29–32, 2019. ISSN 1532-6535. doi: 10.1002/cpt.1275. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.1275. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpt.1275.

Andrea Gaedigk, Scott T Casey, Michelle Whirl-Carrillo, Neil A Miller, and Teri E Klein. PharmVar: A Global Resource and Repository for Pharmacogene Variation. *Clinical pharmacology and therapeutics*, 110(3):542–545, September 2021. ISSN 0009-9236. doi: 10.1002/cpt.2321. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8725060/.

Lukas Gerasimavicius, Sarah A. Teichmann, and Joseph A. Marsh. Leveraging protein structural information to improve variant effect prediction. *Current Opinion in Structural Biology*, 92:103023, June 2025. ISSN 0959-440X. doi: 10.1016/j.sbi.2025.103023. URL https://www.sciencedirect.com/science/article/pii/S0959440X25000417.

Anton Gerilovych, M. Sushko, S. Mandyhra, N. Rodyna, M.Ye Romanko, M. Kuchinskiy, and I. Gerilovych. APPLICATION OF THE NEXT GENERATION SEQUENCING IN BIOLOGY AND MEDICINE. *One Health Journal*, 2:32–44, March 2024. doi: 10.31073/onehealthjournal2024-I-05.

Brian Gilani and Manouchkathe Cassagnol. Biochemistry, Cytochrome P450. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2025. URL http://www.ncbi.nlm.nih.gov/books/NBK557698/.

Jonas A. Gustafson, Sophia B. Gibson, Nikhita Damaraju, Miranda P.G. Zalusky, Kendra Hoekzema, David Twesigomwe, Lei Yang, Anthony A. Snead, Phillip A. Richmond, Wouter De Coster, Nathan D. Olson, Andrea Guarracino, Qiuhui Li, Angela L. Miller, Joy Goffena, Zachary B. Anderson, Sophie H.R. Storz, Sydney A. Ward, Maisha Sinha, Claudia Gonzaga-Jauregui, Wayne E. Clarke, Anna O. Basile, André Corvelo, Catherine Reeves, Adrienne Helland, Rajeeva Lochan Musunuri, Mahler Revsine, Karynne E. Patterson, Cate R. Paschal, Christina Zakarian, Sara Goodwin, Tanner D. Jensen, Esther Robb, The 1000 Genomes ONT Sequencing Consortium, University of Washington Center for Rare Disease Research (UW-CRDR), Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium, William Richard McCombie, Fritz J. Sedlazeck, Justin M. Zook, Stephen B. Montgomery, Erik Garrison, Mikhail Kolmogorov, Michael C. Schatz, Richard N. McLaughlin, Harriet Dashnow, Michael C. Zody, Matt Loose, Miten Jain, Evan E. Eichler, and Danny E. Miller. High-coverage nanopore sequencing of samples from the 1000 Genomes Project to build a comprehensive catalog of human genetic variation. *Genome Research*, 34(11):2061–2073, November 2024. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.279273.124. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.279273.124.

Suhad Hafidh, Hayder Abdul-Amir Makki Al-Hindy, Hussein OM Al-Dahmoshi, Noor SK Al-Khafaji, and Auday Majeed. Insights on Pharmacogenetics and Pharmacogenomics:

Advantages and Challenges for Healthcare Professionals. *Jour Med Resh and Health Sci*, 6 (9):3732–2741, September 2023. ISSN 2589-9031. doi: 10.52845/JMRHS/2023-6-9-3. URL https://jmrhs.info/index.php/jmrhs/article/view/745. Number: 9.

Susanne B. Haga and Nancy M. Allen LaPointe. The Potential Impact of Pharmacogenetic Testing on Medication Adherence. *The pharmacogenomics journal*, 13(6):481–483, December 2013. ISSN 1470-269X. doi: 10.1038/tpj.2013.33. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969027/.

Ananth Hari, Qinghui Zhou, Nina Gonzaludo, John Harting, Stuart A. Scott, Xiang Qin, Steve Scherer, S. Cenk Sahinalp, and Ibrahim Numanagić. An efficient genotyper and star-allele caller for pharmacogenomics. *Genome Research*, 33(1):61–70, January 2023. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.277075.122. URL http://genome.cshlp.org/content/33/1/61. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

Peter W Harrison, M Ridwan Amode, Olanrewaju Austine-Orimoloye, Andrey G Azov, Matthieu Barba, If Barnes, Arne Becker, Ruth Bennett, Andrew Berry, Jyothish Bhai, and others. Ensembl 2024. *Nucleic acids research*, 52(D1):D891–D899, 2024. Publisher: Oxford University Press.

Marwa S. Hassan, A. A. Shaalan, M. I. Dessouky, Abdelaziz E. Abdelnaiem, and Mahmoud ElHefnawi. Evaluation of computational techniques for predicting non-synonymous single nucleotide variants pathogenicity. *Genomics*, 111(4):869–882, July 2019. ISSN 0888-7543. doi: 10.1016/j.ygeno.2018.05.013. URL https://www.sciencedirect.com/science/article/pii/S0888754318301800.

Wei He, Felix Grassmann, Mikael Eriksson, Erik Eliasson, Sara Margolin, Linda Thorén, Per Hall, and Kamila Czene. CYP2D6 Genotype Predicts Tamoxifen Discontinuation and Prognosis in Patients With Breast Cancer. *Journal of Clinical Oncology*, 38(6):548–557, February 2020. ISSN 0732-183X. doi: 10.1200/JCO.19.01535. URL https://ascopubs.org/doi/10.1200/JCO.19.01535. Publisher: Wolters Kluwer.

Linda M. Henricks, Carin A. T. C. Lunenburg, Femke M. de Man, Didier Meulendijks, Geert W. J. Frederix, Emma Kienhuis, Geert-Jan Creemers, Arnold Baars, Vincent O. Dezentjé, Alexander L. T. Imholz, Frank J. F. Jeurissen, Johanna E. A. Portielje, Rob L. H. Jansen, Paul Hamberg, Albert J. ten Tije, Helga J. Droogendijk, Miriam Koopman, Peter Nieboer, Marlène H. W. van de Poel, Caroline M. P. W. Mandigers, Hilde Rosing, Jos H. Beijnen, Erik van Werkhoven, André B. P. van Kuilenburg, Ron H. N. van Schaik, Ron H. J. Mathijssen, Jesse J. Swen, Hans Gelderblom, Annemieke Cats, Henk-Jan Guchelaar, and Jan H. M. Schellens. DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *The Lancet Oncology*, 19(11):1459–1467, November 2018. ISSN 1470-2045, 1474-5488. doi:

10.1016/S1470-2045(18)30686-7. URL https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(18)30686-7/fulltext. Publisher: Elsevier.

Sarah E. Hunt, Benjamin Moore, Ridwan M. Amode, Irina M. Armean, Diana Lemos, Aleena Mushtaq, Andrew Parton, Helen Schuilenburg, Michał Szpak, Anja Thormann, Emily Perry, Stephen J. Trevanion, Paul Flicek, Andrew D. Yates, and Fiona Cunningham. Annotating and prioritizing genomic variants using the Ensembl Variant Effect Predictor—A tutorial. *Human Mutation*, 43(8):986–997, 2022. ISSN 1098-1004. doi: 10.1002/humu.24298. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/humu.24298. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/humu.24298.

Angie C. Jelin and Neeta Vora. Whole exome sequencing: Applications in Prenatal Genetics. *Obstetrics and gynecology clinics of North America*, 45(1):69–81, March 2018. ISSN 0889-8545. doi: 10.1016/j.ogc.2017.10.003. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5813701/.

Yuan Ji and Sherin Shaaban. Interrogating Pharmacogenetics Using Next-Generation Sequencing. *The Journal of Applied Laboratory Medicine*, 9(1):50–60, January 2024. ISSN 2475-7241. doi: 10.1093/jalm/jfad097. URL https://doi.org/10.1093/jalm/jfad097.

Ja Johnson, Ke Caudle, L Gong, M Whirl-Carrillo, Cm Stein, Sa Scott, Mt Lee, Bf Gage, Se Kimmel, Ma Perera, Jl Anderson, M Pirmohamed, Te Klein, Na Limdi, Lh Cavallari, and M Wadelius. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update. *Clinical Pharmacology & Therapeutics*, 102(3):397–404, September 2017. ISSN 0009-9236, 1532-6535. doi: 10.1002/cpt.668. URL https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.668. Publisher: Wiley.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596 (7873):583–589, August 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03819-2. URL https://www.nature.com/articles/s41586-021-03819-2. Publisher: Nature Publishing Group.

Katja S. Just, Richard M. Turner, Vita Dolžan, Erika Cecchin, Jesse J. Swen, David Gurwitz, and Julia C. Stingl. Educating the Next Generation of Pharmacogenomics Experts: Global Educational Needs and Concepts. *Clinical Pharmacology and Therapeutics*, 106(2):313–316, August 2019. ISSN 0009-9236. doi: 10.1002/cpt.1471. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6771464/.

Danya Kabbani, Reem Akika, Ahmed Wahid, Ann K Daly, Ingolf Cascorbi, and Nathalie Khoueiry Zgheib. Pharmacogenomics in practice: a review and implementation guide. *Frontiers in Pharmacology*, 14:1189976, 2023. Publisher: Frontiers Media SA.

W. Kalow. Pharmacogenetics and pharmacogenomics: origin, status, and the hope for personalized medicine. *The Pharmacogenomics Journal*, 6(3):162–165, May 2006. ISSN 1473-1150. doi: 10.1038/sj.tpj.6500361. URL https://www.nature.com/articles/6500361. Publisher: Nature Publishing Group.

Vasudev Kantae, Elke H. J. Krekels, Michiel J. Van Esdonk, Peter Lindenburg, Amy C. Harms, Catherijne A. J. Knibbe, Piet H. Van der Graaf, and Thomas Hankemeier. Integration of pharmacometabolomics with pharmacokinetics and pharmacodynamics: towards personalized drug therapy. *Metabolomics*, 13(1):9, December 2016. ISSN 1573-3890. doi: 10.1007/s11306-016-1143-1. URL https://doi.org/10.1007/s11306-016-1143-1.

Pramod Katara and Anamika Yadav. Pharmacogenes (PGx-genes): Current understanding and future directions. *Gene*, 718:144050, November 2019. ISSN 03781119. doi: 10.1016/j.gene.2019.144050. URL https://linkinghub.elsevier.com/retrieve/pii/S0378111919307097.

Martin Kircher, Daniela M. Witten, Preti Jain, Brian J. O'Roak, Gregory M. Cooper, and Jay Shendure. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, 46(3):310–315, March 2014. ISSN 1546-1718. doi: 10.1038/ng.2892. URL https://www.nature.com/articles/ng.2892. Publisher: Nature Publishing Group.

Merve Nur Koroglu and Kaya Bilguvar. A Comprehensive Bioinformatics Approach to Analysis of Variants: Variant Calling, Annotation, and Prioritization. In Bony De Kumar and Ashish Shelar, editors, *HOX Genes: Methods and Protocols*, pages 207–233. Springer US, New York, NY, 2025. ISBN 978-1-0716-4322-8. doi: 10.1007/978-1-0716-4322-8_15. URL https://doi.org/10.1007/978-1-0716-4322-8_15.

Prateek Kumar, Steven Henikoff, and Pauline C Ng. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4 (7):1073–1081, July 2009. ISSN 1754-2189, 1750-2799. doi: 10.1038/nprot.2009.86. URL https://www.nature.com/articles/nprot.2009.86.

Y. W. Francis Lam and Larisa H. Cavallari. Chapter 1 - Principles of Pharmacogenomics: Pharmacokinetic, Pharmacodynamic, and Clinical Implications. In Yui-Wing Francis Lam and Larisa H. Cavallari, editors, *Pharmacogenomics*, pages 1–44. Academic Press, San Diego, January 2013. ISBN 978-0-12-391918-2. doi: 10.1016/B978-0-12-391918-2.00001-9. URL https://www.sciencedirect.com/science/article/pii/B9780123919182000019.

Immaculate M. Langmia, Katja S. Just, Sabrina Yamoune, Jürgen Brockmöller, Collen Masimirembwa, and Julia C. Stingl. CYP2B6 Functional Variability in Drug Metabolism and Exposure Across Populations—Implication for Drug Safety, Dosing, and Individualized Therapy. *Frontiers in Genetics*, 12:692234, July 2021. ISSN 1664-8021. doi: 10.3389/fgene.2021.692234. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8313315/.

Ming Ta Michael Lee and Teri E Klein. Pharmacogenetics of warfarin: challenges and opportunities. *Journal of Human Genetics*, 58(6):334–338, June 2013. ISSN 1434-5161, 1435-232X. doi: 10.1038/jhg.2013.40. URL https://www.nature.com/articles/jhg201340. Publisher: Springer Science and Business Media LLC.

Seung-been Lee, Marsha M. Wheeler, Karynne Patterson, Sean McGee, Rachel Dalton, Erica L. Woodahl, Andrea Gaedigk, Kenneth E. Thummel, and Deborah A. Nickerson. Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. *Genetics in medicine : official journal of the American College of Medical Genetics*, 21(2):361–372, February 2019a. ISSN 1098-3600. doi: 10.1038/s41436-018-0054-0. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6281872/.

Seung-been Lee, Marsha M. Wheeler, Kenneth E. Thummel, and Deborah A. Nickerson. Calling Star Alleles With Stargazer in 28 Pharmacogenes With Whole Genome Sequences. *Clinical Pharmacology & Therapeutics*, 106(6):1328–1337, 2019b. ISSN 1532-6535. doi: 10.1002/cpt.1552. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.1552. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpt.1552.

Yuhua Li, Qiang Meng, Mengbi Yang, Dongyang Liu, Xiangyu Hou, Lan Tang, Xin Wang, Yuanfeng Lyu, Xiaoyan Chen, Kexin Liu, Ai-Ming Yu, Zhong Zuo, and Huichang Bi. Current trends in drug metabolism and pharmacokinetics. *Acta Pharmaceutica Sinica B*, 9(6):1113–1144, November 2019. ISSN 2211-3835. doi: 10.1016/j.apsb.2019.10.001. URL https://www.sciencedirect.com/science/article/pii/S2211383519304800.

Ye Liu, William S. B. Yeung, Philip C. N. Chiu, and Dandan Cao. Computational approaches for predicting variant impact: An overview from resources, principles to applications. *Frontiers in Genetics*, 13, September 2022. ISSN 1664-8021. doi: 10.3389/fgene.2022.981005. URL https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.981005/full. Publisher: Frontiers.

Lonah, Zita Arieselia, Jonny Setiawan, Rita Dewi, Maria Dara Novi Handayani, and Linawati Hananta. Pharmacogenomic Role in Cytochrome P450 Enzymes-Mediated Metabolism for Drug Therapy. *Journal of Urban Health Research*, 1(3):44–56, August 2023. ISSN 2964-4194. doi: 10.25170/juhr.v1i3.4629. URL https://ejournal.atmajaya.ac.id/index.php/juhr/article/view/4629. Number: 3.

Jasmine A. Luzum, Natasha Petry, Annette K. Taylor, Sara L. Van Driest, Henry M. Dunnenberger, and Larisa H. Cavallari. Moving Pharmacogenetics into Practice: It's All About the Evidence! *Clinical pharmacology and therapeutics*, 110(3):649–661, September

2021. ISSN 0009-9236. doi: 10.1002/cpt.2327. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8376790/.

Leen Magarbeh, Ilona Gorbovskaya, Bernard Le Foll, Reuven Jhirad, and Daniel J. Müller. Reviewing pharmacogenetics to advance precision medicine for opioids. *Biomedicine & Pharmacotherapy*, 142:112060, October 2021. ISSN 07533322. doi: 10.1016/j.biopha.2021.112060. URL https://linkinghub.elsevier.com/retrieve/pii/S075333222100843X.

Khalid Mahmood, Chol-hee Jung, Gayle Philip, Peter Georgeson, Jessica Chung, Bernard J Pope, and Daniel J Park. Variant effect prediction tools assessed using independent, functional assay-based datasets: implications for discovery and diagnostics. *Human genomics*, 11:1–8, 2017. Publisher: Springer.

Lauren Marcath, Allison M Deal, Emily Van Wieren, William Danko, Christine M Walko, Joseph G Ibrahim, Karen E Weck, David R Jones, Zeruesenay Desta, Howard L McLeod, Lisa A Carey, William J Irvin, and Daniel L Hertz. Comprehensive assessment of cytochromes P450 and transporter genetics with endoxifen concentration during tamoxifen treatment. *Pharmacogenetics and genomics*, 27(11):402–409, November 2017. ISSN 1744-6872. doi: 10.1097/FPC.0000000000000311. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5659294/.

Abdullah Al Maruf and Chad A. Bousman. Approaches and hurdles of implementing pharmacogenetic testing in the psychiatric clinic. *Psychiatry and Clinical Neurosciences Reports*, 1(2):e26, 2022. ISSN 2769-2558. doi: 10.1002/pcn5.26. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pcn5.26. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pcn5.26.

Philip J Mason, José M Bautista, and Florinda Gilsanz. G6PD deficiency: the genotype-phenotype association. *Blood Reviews*, 21(5):267–283, September 2007. ISSN 0268-960X. doi: 10.1016/j.blre.2007.05.002. URL https://www.sciencedirect.com/science/article/pii/S0268960X0700032X.

Maja Matic, Marga Nijenhuis, Bianca Soree, Nienke J. de Boer-Veger, Anne-Marie Buunk, Elisa J. F. Houwink, Hans Mulder, Gerard A. P. J. M. Rongen, Jan van der Weide, Bob Wilffert, Jesse J. Swen, Henk-Jan Guchelaar, Vera H. M. Deneer, and Ron H. N. van Schaik. Dutch Pharmacogenetics Working Group (DPWG) guideline for the gene–drug interaction between CYP2D6 and opioids (codeine, tramadol and oxycodone). *European Journal of Human Genetics*, 30(10):1105–1113, October 2022. ISSN 1476-5438. doi: 10.1038/s41431-021-00920-y. URL https://www.nature.com/articles/s41431-021-00920-y. Publisher: Nature Publishing Group.

William McLaren, Laurent Gil, Sarah E. Hunt, Harpreet Singh Riat, Graham R. S. Ritchie, Anja Thormann, Paul Flicek, and Fiona Cunningham. The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1):122, June 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-0974-4. URL https://doi.org/10.1186/s13059-016-0974-4.

John Meletis and Kostas Konstantopoulos. Favism-from the "avoid fava beans" of Pythagoras to the present. *Haema*, 7(1):17–21, 2004.

Urs A. Meyer, Susan G. Amara, Terrence F. Blaschke, and Paul A. Insel. Introduction to the Theme "Pharmacological Individuality: New Insights and Strategies for Personalized and Precise Drug Treatment". *Annual Review of Pharmacology and Toxicology*, 64(Volume 64, 2024):27–31, January 2024. ISSN 0362-1642, 1545-4304. doi: 10.1146/annurev-pharmtox-090123-010552. URL https://www.annualreviews.org/content/journals/10.1146/annurev-pharmtox-090123-010552. Publisher: Annual Reviews.

Clint Mizzi, Brock Peters, Christina Mitropoulou, Konstantinos Mitropoulos, Theodora Katsila, Misha R Agarwal, Ron Hn Van Schaik, Radoje Drmanac, Joseph Borg, and George P Patrinos. Personalized Pharmacogenomics Profiling Using Whole-Genome Sequencing. *Pharmacogenomics*, 15(9):1223–1234, July 2014. ISSN 1462-2416, 1744-8042. doi: 10.2217/pgs.14.102. URL https://www.tandfonline.com/doi/full/10.2217/pgs.14.102.

Sarah A. Morris, Ashraf T. Alsaidi, Allison Verbyla, Adilen Cruz, Casey Macfarlane, Joseph Bauer, and Jai N. Patel. Cost Effectiveness of Pharmacogenetic Testing for Drugs with Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines: A Systematic Review. *Clinical Pharmacology and Therapeutics*, 112(6):1318–1328, December 2022. ISSN 0009-9236. doi: 10.1002/cpt.2754. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9828439/.

Arno G. Motulsky. DRUG REACTIONS, ENZYMES, AND BIOCHEMICAL GENETICS. *Journal of the American Medical Association*, 165(7):835–837, October 1957. ISSN 0002-9955. doi: 10.1001/jama.1957.72980250010016. URL https://doi.org/10.1001/jama.1957.72980250010016.

Tessa A. M. Mulder, Mirjam de With, Marzia del Re, Romano Danesi, Ron H. J. Mathijssen, and Ron H. N. van Schaik. Clinical CYP2D6 Genotyping to Personalize Adjuvant Tamoxifen Treatment in ER-Positive Breast Cancer Patients: Current Status of a Controversy. *Cancers*, 13(4):771, January 2021. ISSN 2072-6694. doi: 10.3390/cancers13040771. URL https://www.mdpi.com/2072-6694/13/4/771. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

Monica Neamțu. MOLECULAR BACKGROUNDS OF INDIVIDUALISED PHARMACOTHERAPY. *FARMACIA*, 68(3):396–405, June 2020. ISSN 00148237. doi: 10.31925/farmacia.2020.3.4. URL http://farmaciajournal.com/issue-articles/molecular-backgrounds-of-individualised-pharmacotherapy/.

Pauline C. Ng and Steven Henikoff. Predicting Deleterious Amino Acid Substitutions. *Genome Research*, 11(5):863–874, May 2001. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.176601. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.176601.

Wayne T. Nicholson, Christine M. Formea, Eric T. Matey, Jessica A. Wright, Jyothsna Giri, and Ann M. Moyer. Considerations When Applying Pharmacogenomics to Your Practice. *Mayo Clinic Proceedings*, 96(1):218–230, January 2021. ISSN 00256196. doi: 10.1016/j.mayocp.2020.03.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0025619620302688.

Ella T. Nkhoma, Charles Poole, Vani Vannappagari, Susan A. Hall, and Ernest Beutler. The global prevalence of glucose-6-phosphate dehydrogenase deficiency: A systematic review and meta-analysis. *Blood Cells, Molecules, and Diseases*, 42(3):267–278, May 2009. ISSN 1079-9796. doi: 10.1016/j.bcmd.2008.12.005. URL https://www.sciencedirect.com/science/article/pii/S1079979609000059.

Charity Nofziger, Amy J. Turner, Katrin Sangkuhl, Michelle Whirl-Carrillo, José A. G. Agúndez, John L. Black, Henry M. Dunnenberger, Gualberto Ruano, Martin A. Kennedy, Michael S. Phillips, Houda Hachad, Teri E. Klein, and Andrea Gaedigk. PharmVar GeneReview: CYP2D6. *Clinical pharmacology and therapeutics*, 107(1):154–170, January 2020. ISSN 0009-9236. doi: 10.1002/cpt.1643. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6925641/.

Silvia Novello, Marina Longo, and Matteo Giaj Levra. Toward Therapies Tailored to Patient Characteristics. *Journal of Thoracic Oncology*, 2(5):S38–S41, May 2007. ISSN 1556-0864, 1556-1380. doi: 10.1097/01.JTO.0000268642.40826.98. URL https://www.jto.org/article/S1556-0864(15)31398-8/fulltext. Publisher: Elsevier.

Rostam Osanlou, Lauren Walker, Dyfrig A Hughes, Girvan Burnside, and Munir Pirmohamed. Adverse drug reactions, multimorbidity and polypharmacy: a prospective analysis of 1 month of medical admissions. *BMJ Open*, 12(7):e055551, July 2022. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2021-055551. URL https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2021-055551.

Ryan P. Owen, Katrin Sangkuhl, Teri E. Klein, and Russ B. Altman. Cytochrome P450 2D6. *Pharmacogenetics and Genomics*, 19(7):559, July 2009. ISSN 1744-6872. doi: 10.1097/FPC.0b013e32832e0e97. URL https://journals.lww.com/jpharmacogenetics/citation/2009/07000/cytochrome_p450_2d6.10.aspx.

Maria-Theodora Pandi, Maria Koromina, Iordanis Tsafaridis, Sotirios Patsilinakos, Evangelos Christoforou, Peter J. van der Spek, and George P. Patrinos. A novel machine learning-based approach for the computational functional assessment of pharmacogenomic variants. *Human Genomics*, 15(1):51, August 2021. ISSN 1479-7364. doi: 10.1186/s40246-021-00352-1. URL https://doi.org/10.1186/s40246-021-00352-1.

Elena Peruzzi, Rossana Roncato, Elena De Mattia, Alessia Bignucolo, Jesse J. Swen, Henk-Jan Guchelaar, Giuseppe Toffoli, and Erika Cecchin. Implementation of pre-emptive testing of a pharmacogenomic panel in clinical practice: Where do we stand? *British Journal of Clinical Pharmacology*, 91(2):270–282, 2025. ISSN 1365-2125. doi: 10.1111/bcp.

15956. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/bcp.15956. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bcp.15956.

Munir Pirmohamed. Pharmacogenetics: past, present and future. *Drug Discovery Today*, 16 (19-20):852–861, October 2011. ISSN 13596446. doi: 10.1016/j.drudis.2011.08.006. URL https://linkinghub.elsevier.com/retrieve/pii/S1359644611002571.

Munir Pirmohamed. Pharmacogenomics: current status and future perspectives. *Nature Reviews Genetics*, 24(6):350–362, June 2023. ISSN 1471-0064. doi: 10.1038/s41576-022-00572-8. URL https://www.nature.com/articles/s41576-022-00572-8. Publisher: Nature Publishing Group.

V. Ramensky. Human non-synonymous SNPs: server and survey. *Nucleic Acids Research*, 30 (17):3894–3900, September 2002. ISSN 13624962. doi: 10.1093/nar/gkf493. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkf493.

Andrijana Rebekić, Zdenko Lončarić, Sonja Petrović, and Sonja Marić. Pearson's or Spearman's correlation coefficient-which one to use? *Poljoprivreda*, 21(2):47–54, 2015.

Catherine Rehder, Lora J. H. Bean, David Bick, Elizabeth Chao, Wendy Chung, Soma Das, Julianne O'Daniel, Heidi Rehm, Vandana Shashi, and Lisa M. Vincent. Next-generation sequencing for constitutional variants in the clinical laboratory, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 23(8):1399–1415, August 2021. ISSN 1098-3600, 1530-0366. doi: 10.1038/s41436-021-01139-4. URL https://www.gimjournal.org/article/S1098-3600(21)05056-5/fulltext. Publisher: Elsevier.

Mary V. Relling, Matthias Schwab, Michelle Whirl-Carrillo, Guilherme Suarez-Kurtz, Ching-Hon Pui, Charles M. Stein, Ann M. Moyer, William E. Evans, Teri E. Klein, Federico Guillermo Antillon-Klussmann, Kelly E. Caudle, Motohiro Kato, Allen E.J. Yeoh, Kjeld Schmiegelow, and Jun J. Yang. Clinical Pharmacogenetics Implementation Consortium Guideline for Thiopurine Dosing Based on <*span style="font-variant:small-caps;"*>*TPMT*</*span*> and <*span style="font-variant:small-caps;"*>*NUDT*</*span*>*15* Genotypes: 2018 Update. *Clinical Pharmacology & Therapeutics*, 105(5):1095–1105, May 2019. ISSN 0009-9236, 1532-6535. doi: 10.1002/cpt.1304. URL https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.1304. Publisher: Wiley.

MV Relling and TE Klein. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clinical Pharmacology and Therapeutics*, 89 (3):464–467, March 2011. ISSN 0009-9236. doi: 10.1038/clpt.2010.279. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3098762/.

Philipp Rentzsch, Daniela Witten, Gregory M Cooper, Jay Shendure, and Martin Kircher. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1):D886–D894, January 2019. ISSN 0305-1048. doi: 10.1093/nar/gky1016. URL https://doi.org/10.1093/nar/gky1016.

Cristian Riccio, Max L Jansen, Linlin Guo, and Andreas Ziegler. Variant effect predictors: a systematic review and practical guide. *Human Genetics*, 143(5):625–634, 2024. Publisher: Springer.

J. D. Robarge, L. Li, Z. Desta, A. Nguyen, and D. A. Flockhart. The star-allele nomenclature: retooling for translational genomics. *Clinical Pharmacology and Therapeutics*, 82(3):244–248, September 2007. ISSN 0009-9236. doi: 10.1038/sj.clpt.6100284.

Laura E Russell, , and Ute I Schwarz. Variant Discovery Using Next-Generation Sequencing and its Future Role in Pharmacogenetics. *Pharmacogenomics*, 21(7):471–486, May 2020. ISSN 1462-2416. doi: 10.2217/pgs-2019-0190. URL https://doi.org/10.2217/pgs-2019-0190. Publisher: Taylor & Francis _eprint: https://doi.org/10.2217/pgs-2019-0190.

Simeon Rüdesheim, Dominik Selzer, Thomas Mürdter, Svitlana Igel, Reinhold Kerb, Matthias Schwab, and Thorsten Lehr. Physiologically Based Pharmacokinetic Modeling to Describe the CYP2D6 Activity Score-Dependent Metabolism of Paroxetine, Atomoxetine and Risperidone. *Pharmaceutics*, 14(8):1734, August 2022. ISSN 1999-4923. doi: 10.3390/pharmaceutics14081734. URL https://www.mdpi.com/1999-4923/14/8/1734. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

Aga Syed Sameer, Mujeeb Zafar Banday, and Saniya Nissar. Mutations and Polymorphisms: What Is The Difference? In Aga Syed Sameer, Mujeeb Zafar Banday, and Saniya Nissar, editors, *Genetic Polymorphism and cancer susceptibility*, pages 1–21. Springer, Singapore, 2021. ISBN 978-981-336-699-2. doi: 10.1007/978-981-33-6699-2_1. URL https://doi.org/10.1007/978-981-33-6699-2_1.

F. Sanger, S. Nicklen, and A. R. Coulson. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467, December 1977. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.74.12.5463. URL https://pnas.org/doi/full/10.1073/pnas.74.12.5463.

Heena Satam, Kandarp Joshi, Upasana Mangrolia, Sanober Waghoo, Gulnaz Zaidi, Shravani Rawool, Ritesh P. Thakare, Shahid Banday, Alok K. Mishra, Gautam Das, and Sunil K. Malonia. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*, 12(7):997, July 2023. ISSN 2079-7737. doi: 10.3390/biology12070997. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10376292/.

Max Schubach, Thorben Maass, Lusiné Nazaretyan, Sebastian Röner, and Martin Kircher. CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. *Nucleic Acids Research*, 52(D1): D1143–D1154, January 2024. ISSN 0305-1048. doi: 10.1093/nar/gkad989. URL https://doi.org/10.1093/nar/gkad989.

Ute I. Schwarz, Markus Gulilat, and Richard B. Kim. The Role of Next-Generation Sequencing in Pharmacogenetics and Pharmacogenomics. *Cold Spring Harbor Perspectives in Medicine*, 9(2):a033027, February 2019. ISSN , 2157-1422. doi: 10.1101/cshperspect.a033027. URL http://perspectivesinmedicine.cshlp.org/content/9/2/a033027. Publisher: Cold Spring Harbor Laboratory Press.

Mital Shah, Shanks , Morag, Packham , Emily, Williams , Jonathan, Haysmoore , Jesse, MacLaren , Robert E., Németh , Andrea H., Clouston , Penny, , and Susan M. Downes. Next generation sequencing using phenotype-based panels for genetic testing in inherited retinal diseases. *Ophthalmic Genetics*, 41(4):331–337, July 2020. ISSN 1381-6810. doi: 10.1080/13816810.2020.1778736. URL https://doi.org/10.1080/13816810.2020.1778736. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13816810.2020.1778736.

GiWon Shin, Stephanie U Greer, Li C Xia, HoJoon Lee, Jun Zhou, T Christian Boles, and Hanlee P Ji. Targeted short read sequencing and assembly of re-arrangements and candidate gene loci provide megabase diplotypes. *Nucleic Acids Research*, 47(19):e115–e115, November 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkz661. URL https://academic.oup.com/nar/article/47/19/e115/5539887.

Maria Shoaib, Mansoor Ali Merchant Rameez, Syed Ather Hussain, Mohammed Madadin, and Ritesh G. Menezes. Personalized Medicine in a New Genomic Era: Ethical and Legal Aspects. *Science and Engineering Ethics*, 23(4):1207–1212, August 2017. ISSN 1471-5546. doi: 10.1007/s11948-016-9828-4. URL https://doi.org/10.1007/s11948-016-9828-4.

Cesare R. Sirtori. The pharmacology of statins. *Pharmacological Research*, 88:3–11, October 2014. ISSN 1043-6618. doi: 10.1016/j.phrs.2014.03.002. URL https://www.sciencedirect.com/science/article/pii/S1043661814000255.

Johanna Sistonen, Antti Sajantila, Oscar Lao, Jukka Corander, Guido Barbujani, and Silvia Fuselli. CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. *Pharmacogenetics and Genomics*, 17(2):93, February 2007. ISSN 1744-6872. doi: 10.1097/01.fpc.0000239974.69464.f2. URL https://journals.lww.com/jpharmacogenetics/abstract/2007/02000/cyp2d6_worldwide_genetic_variation_shows_high.1.aspx.

Andrew Somogy. Evolution of pharmacogenomics. *Proceedings of the Western Pharmacology Society*, 51:1–4, 2008. ISSN 0083-8969.

Julius Olugbenga Soyinka, Charles Okeke Nnadi, and Cyprian Ogbonna Onyeji. Insights and Current Perspectives on Pharmacogenomics of Antimalarial Drugs, September 2022. URL https://www.researchsquare.com/article/rs-2030964/v1.

Srilakshmi Srinivasan, Clements , Judith A., , and Jyotsna Batra. Single nucleotide polymorphisms in clinics: Fantasy or reality for cancer? *Critical Reviews in Clinical Laboratory Sciences*, 53(1):29–39, January 2016. ISSN 1040-8363. doi: 10.3109/10408363.2015.

1075469. URL https://doi.org/10.3109/10408363.2015.1075469. Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/10408363.2015.1075469.

DIMITRIOS A Stakos and Harisios Boudoulas. Pharmacogenetics and pharmacogenomics in cardiology. *Hellenic J Cardiol*, 43:1–15, 2002.

Marlaina R. Stocco and Rachel F. Tyndale. Chapter Three - Cytochrome P450 enzymes and metabolism of drugs and neurotoxins within the mammalian brain. In Hiroshi Yamazaki, editor, *Advances in Pharmacology*, volume 95 of *Pharmacology and Toxicology of Cytochrome P450 – 60th Anniversary*, pages 73–106. Academic Press, January 2022. doi: 10.1016/bs.apha.2022.04.003. URL https://www.sciencedirect.com/science/article/pii/S1054358922000278.

Anita Stojanović Marković, Matea Zajc Petranović, Tatjana Škarić Jurić, Željka Celinščak, Maja Šetinc, Željka Tomas, and Marijana Peričić Salihović. Relevance of CYP2D6 Gene Variants in Population Genetic Differentiation. *Pharmaceutics*, 14(11):2481, November 2022. ISSN 1999-4923. doi: 10.3390/pharmaceutics14112481. URL https://www.mdpi.com/1999-4923/14/11/2481. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

Aneesh T P, Sonal Sekhar M, Asha Jose, Lekshmi Chandran, and Subin Mary Zachariah. Pharmacogenomics: The Right Drug to the Right Person. *Journal of Clinical Medicine Research*, 1(4):191–194, October 2009. ISSN 1918-3003. doi: 10.4021/jocmr2009.08.1255. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3299179/.

Alireza Tafazoli, Henk-Jan Guchelaar, Wojciech Miltyk, Adam J. Kretowski, and Jesse J. Swen. Applying Next-Generation Sequencing Platforms for Pharmacogenomic Testing in Clinical Practice. *Frontiers in Pharmacology*, 12, August 2021. ISSN 1663-9812. doi: 10.3389/fphar.2021.693453. URL https://www.frontiersin.orghttps://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2021.693453/full. Publisher: Frontiers.

Christopher Taylor, Ian Crosby, Vincent Yip, Peter Maguire, Munir Pirmohamed, and Richard M. Turner. A Review of the Important Role of CYP2D6 in Pharmacogenomics. *Genes*, 11(11):1295, November 2020. ISSN 2073-4425. doi: 10.3390/genes11111295. URL https://www.mdpi.com/2073-4425/11/11/1295. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

Roman Tremmel, Sebastian Pirmann, Yitian Zhou, and Volker M. Lauschke. Translating pharmacogenomic sequencing data into drug response predictions—How to interpret variants of unknown significance. *British Journal of Clinical Pharmacology*, 91(2):252–263, 2025. ISSN 1365-2125. doi: 10.1111/bcp.15915. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/bcp.15915. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bcp.15915.

Rachel J. Tyson, Christine C. Park, J. Robert Powell, J. Herbert Patterson, Daniel Weiner, Paul B. Watkins, and Daniel Gonzalez. Precision Dosing Priority Criteria: Drug, Disease, and Patient Population Variables. *Frontiers in Pharmacology*, 11, April 2020. ISSN 1663-9812. doi: 10.3389/fphar.2020.00420. URL https://www.frontiersin.orghttps://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2020.00420/full. Publisher: Frontiers.

Denise van der Drift, Mirjam Simoons, Birgit C. P. Koch, Gemma Brufau, Patrick Bindels, Maja Matic, and Ron H. N. van Schaik. Implementation of Pharmacogenetics in First-Line Care: Evaluation of Its Use by General Practitioners. *Genes*, 14(10):1841, September 2023. ISSN 2073-4425. doi: 10.3390/genes14101841. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10606701/.

Maaike van der Lee, Marjolein Kriek, Henk-Jan Guchelaar, and Jesse J. Swen. Technologies for Pharmacogenomics: A Review. *Genes*, 11(12):1456, December 2020. ISSN 2073-4425. doi: 10.3390/genes11121456. URL https://www.mdpi.com/2073-4425/11/12/1456. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

Mihaly Varadi, Damian Bertoni, Paulyna Magana, Urmila Paramval, Ivanna Pidruchna, Malarvizhi Radhakrishnan, Maxim Tsenkov, Sreenath Nair, Milot Mirdita, Jingi Yeo, Oleg Kovalevskiy, Kathryn Tunyasuvunakool, Agata Laydon, Augustin Žídek, Hamish Tomlinson, Dhavanthi Hariharan, Josh Abrahamson, Tim Green, John Jumper, Ewan Birney, Martin Steinegger, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, 52(D1):D368–D375, January 2024. ISSN 0305-1048. doi: 10.1093/nar/gkad1011. URL https://doi.org/10.1093/nar/gkad1011.

Robert Vaser, Swarnaseetha Adusumalli, Sim Ngak Leng, Mile Sikic, and Pauline C. Ng. SIFT missense predictions for genomes. *Nature Protocols*, 11(1):1–9, January 2016. ISSN 1750-2799. doi: 10.1038/nprot.2015.123. URL https://www.nature.com/articles/nprot.2015.123. Publisher: Nature Publishing Group.

Friedrich Vogel. Moderne Probleme der Humangenetik. In L. Heilmeyer, R. Schoen, and B. de Rudder, editors, *Ergebnisse der Inneren Medizin und Kinderheilkunde*, pages 52–125, Berlin, Heidelberg, 1959. Springer. ISBN 978-3-642-94744-5. doi: 10.1007/978-3-642-94744-5_2.

Danyi Wang, Bolleddula , Jayaprakasam, Coenen-Stass , Anna, Grombacher , Thomas, Dong , Jennifer Q, Scheuenpflug , Juergen, Locatelli , Giuseppe, , and Zheng Feng. Implementation of whole-exome sequencing for pharmacogenomics profiling and exploring its potential clinical utilities. *Pharmacogenomics*, 25(4):197–206, March 2024. ISSN 1462-2416. doi: 10.2217/pgs-2023-0243. URL https://doi.org/10.2217/pgs-2023-0243. Publisher: Taylor & Francis _eprint: https://doi.org/10.2217/pgs-2023-0243.

Dong Wang, Jie Li, Yadong Wang, and Edwin Wang. A comparison on predicting functional impact of genomic variants. *NAR Genomics and Bioinformatics*, 4(1):lqab122, January 2022. ISSN 2631-9268. doi: 10.1093/nargab/lqab122. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8759571/.

Amanda Warr, Christelle Robert, David Hume, Alan Archibald, Nader Deeb, and Mick Watson. Exome Sequencing: Current and Future Perspectives. *G3 Genes | Genomes | Genetics*, 5(8):1543–1550, August 2015. ISSN 2160-1836. doi: 10.1534/g3.115.018564. URL https://academic.oup.com/g3journal/article/5/8/1543/6025359.

Richard Weinshilboum and Liewei Wang. Pharmacogenomics: Precision Medicine and Drug Response. *Mayo Clinic proceedings*, 92(11):1711–1722, November 2017. ISSN 0025-6196. doi: 10.1016/j.mayocp.2017.09.001. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5682947/.

Richard M. Weinshilboum and Liewei Wang. Pharmacogenetics and Pharmacogenomics: Development, Science, and Translation. *Annual Review of Genomics and Human Genetics*, 7(Volume 7, 2006):223–245, September 2006. ISSN 1527-8204, 1545-293X. doi: 10.1146/annurev.genom.6.080604.162315. URL https://www.annualreviews.org/content/journals/10.1146/annurev.genom.6.080604.162315. Publisher: Annual Reviews.

Cassandra White, Rodney Scott, Christine L Paul, and Stephen P Ackland. Pharmacogenomics in the era of personalised medicine. *Medical Journal of Australia*, 217(10):510–513, 2022. ISSN 1326-5377. doi: 10.5694/mja2.51759. URL https://onlinelibrary.wiley.com/doi/abs/10.5694/mja2.51759. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.5694/mja2.51759.

Erin D. Williams and Michele M. Schoonmaker. Ethical Issues in Pharmacogenomics. In Dragan Primorac, Wolfgang Höppner, and Lidija Bach-Rojecky, editors, *Pharmacogenomics in Clinical Practice*, pages 347–371. Springer International Publishing, Cham, 2023. ISBN 978-3-031-45903-0. doi: 10.1007/978-3-031-45903-0_19. URL https://doi.org/10.1007/978-3-031-45903-0_19.

Birgit M. Wollmann, Elisabet Størset, Marianne Kristiansen Kringen, Espen Molden, and Robert L. Smith. Prediction of CYP2D6 poor metabolizers by measurements of solanidine and metabolites—a study in 839 patients with known CYP2D6 genotype. *European Journal of Clinical Pharmacology*, 79(4):523–531, April 2023. ISSN 1432-1041. doi: 10.1007/s00228-023-03462-y. URL https://doi.org/10.1007/s00228-023-03462-y.

Ulrich M Zanger, Sebastian Raimundo, and Michel Eichelbaum. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn-Schmiedeberg's archives of pharmacology*, 369:23–37, 2004. Publisher: Springer.

Mingzhe Zhao, Jingsong Ma, Mo Li, Yingtian Zhang, Bixuan Jiang, Xianglong Zhao, Cong Huai, Lu Shen, Na Zhang, Lin He, and Shengying Qin. Cytochrome P450 Enzymes and Drug Metabolism in Humans. *International Journal of Molecular Sciences*, 22 (23):12808, January 2021. ISSN 1422-0067. doi: 10.3390/ijms222312808. URL https://www.mdpi.com/1422-0067/22/23/12808. Number: 23 Publisher: Multidisciplinary Digital Publishing Institute.

Y Zhou, M Ingelman-Sundberg, and Vm Lauschke. Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. *Clinical Pharmacology & Therapeutics*, 102(4):688–700, 2017. ISSN 1532-6535. doi: 10.1002/cpt.690. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.690. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpt.690.

Yitian Zhou and Volker M. Lauschke. Computational Tools to Assess the Functional Consequences of Rare and Noncoding Pharmacogenetic Variability. *Clinical Pharmacology & Therapeutics*, 110(3):626–636, 2021. ISSN 1532-6535. doi: 10.1002/cpt.2289. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.2289. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpt.2289.

Yitian Zhou and Volker M. Lauschke. The genetic landscape of major drug metabolizing cytochrome P450 genes—an updated analysis of population-scale sequencing data. *The Pharmacogenomics Journal*, 22(5-6):284–293, 2022a. ISSN 1470-269X. doi: 10.1038/s41397-022-00288-2. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9674520/.

Yitian Zhou and Volker M. Lauschke. Population pharmacogenomics: an update on ethnogeographic differences and opportunities for precision public health. *Human Genetics*, 141(6):1113–1136, June 2022b. ISSN 1432-1203. doi: 10.1007/s00439-021-02385-x. URL https://doi.org/10.1007/s00439-021-02385-x.

Yitian Zhou, Kohei Fujikura, Souren Mkrtchian, and Volker M. Lauschke. Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data. *Frontiers in Pharmacology*, 9:1437, December 2018. ISSN 1663-9812. doi: 10.3389/fphar.2018.01437. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6288784/.

Yitian Zhou, Carolina Dagli Hernandez, and Volker M. Lauschke. Population-scale predictions of DPD and TPMT phenotypes using a quantitative pharmacogene-specific ensemble classifier. *British Journal of Cancer*, 123(12):1782–1789, December 2020. ISSN 1532-1827. doi: 10.1038/s41416-020-01084-0. URL https://www.nature.com/articles/s41416-020-01084-0. Publisher: Nature Publishing Group.

Yitian Zhou, Roman Tremmel, Elke Schaeffeler, Matthias Schwab, and Volker M. Lauschke. Challenges and opportunities associated with rare-variant pharmacogenomics. *Trends in Pharmacological Sciences*, 43(10):852–865, October 2022. ISSN 01656147. doi: