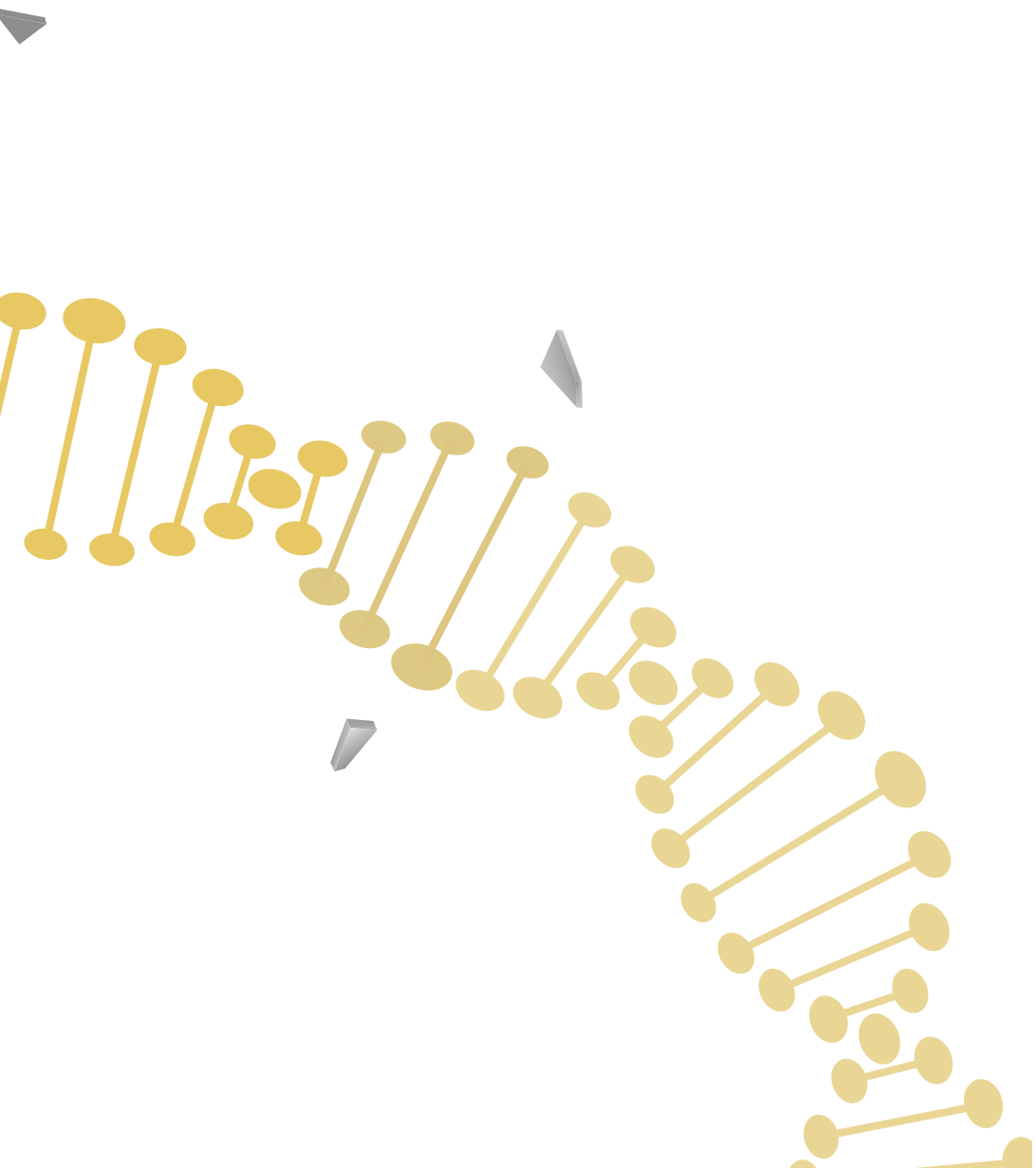
Exploring the role of genetic complexity in pharmacogenomics

Zhai, Q.

Chapter 7

General discussion and future perspectives



Introduction

Pharmacogenomics (PGx) is a propulsive area of personalized medicine which investigates the impact of genetic variability on drug response. A prospective implementation study suggested that genotype-guided treatment significantly reduced the risk of adverse drug reactions with 30% [1]. Importantly, the rapid emergence of novel genome technologies, along with their significant reduction in price, has paved the way towards extensive studies. These studies focus on how genetic variability contributes to inter-individual differences in drug response. However, not all variabilities of drug response can be explained by currently used strategies in PGx. The gap between expected and known genetic contribution on drug response is referred to missing heritability. In this thesis, new approaches were applied to illustrate and partially solve the potential causes for the missing heritability (Figure 7.1). Firstly, long-read sequencing can contribute to resolving complex pharmacogenes, as well as detecting novel and rare variants. Secondly, *in silico* algorithms are required for high-throughput interpretation of variants detected by sequencing. Thirdly, the intricate regulation of pharmacogenes expression, for instance, transcription regulation, has not been studied extensively and can contribute. Lastly, potential new drug-gene associations are explored by genome-wide association study (GWAS).

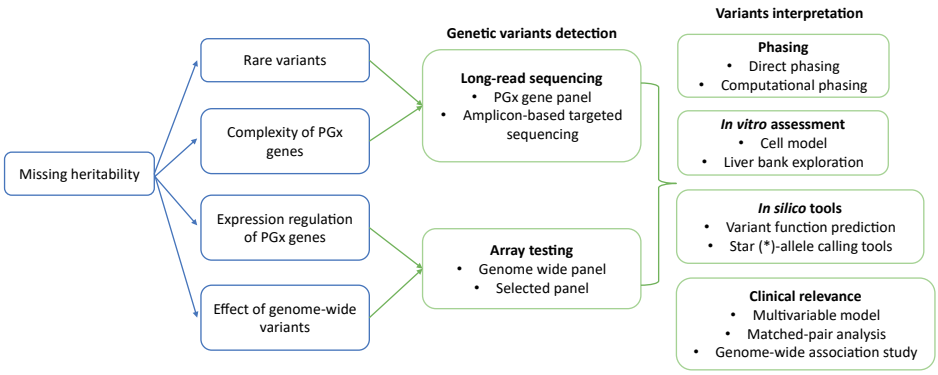


Figure 7.1: Overview of the general discussion.

Discovery of genetics complexity and the contribution of long-read sequencing

Many pharmacogenes are highly complex due to the presence of tandem repeats and segmental duplications, for instance, *CYP2D6* and *CYP2B6* [2, 3]. Nowadays, these complex variants are used in clinics by assigning them to pre-defined *-alleles. Further-

more, some pharmacogenes exhibit high sequence homology with neighboring genes. For example, the four coding genes and pseudogenes in the *CYP3A* locus (*CYP3A4*, *CYP3A5*, *CYP3A7*, *CYP3A43*) share over 90% of their sequence [4]. Long-read sequencing shows promising in resolving these complex regions in pharmacogenes. In **Chapter 3**, we developed a capture-based approach to evaluate the performance of long-read sequencing in PGx. Our long-read approach aims to capture the full length of the locus, covering not only the protein-coding regions of the gene but also noncoding regions such as upstream and downstream regulatory regions and introns. This comprehensive coverage enables us to fully uncover the genetic makeup of the targeted pharmacogenes, facilitating the understanding of their roles in drug response. This approach showed good coverage of most target genes in the panel, including complex regions like *CYP2D6* and the *CYP3A* locus. In 10 selected pharmacogenes (*CYP2D6*, *CYP2C9*, *CYP2C19*, *CYP2B6*, *CYP3A4*, *CYP3A5*, *SLOC1B1*, *VKORC1*, *UGT1A1*, *TPMT*), protein-coding variants account for only a small percentage (about 2%) of total genetic variations. Of these variants, 36% have been assigned to the star(*)-alleles, as they were typically considered deleterious because they lead to amino acid alterations or frameshifts [5]. The assignment from variants to *-allele is a key step for clinical use.

Genetic variants located in the non-coding sequences can play a critical role in gene function. Nonetheless, their clinical impact is still lacking behind [4]. In **Chapter 3**, intronic variants and up-/downstream variants account for 65% and 25% of all variants detected in 10 selected pharmacogenes, respectively. However, less than 2% of these variants are part of the *-allele nomenclature. The mechanisms of how these non-coding variants affect gene function are complex [4, 6]. Firstly, variants in promoters could affect the binding affinities of transcription factors, thereby altering the transcriptional activation. A clinically relevant example of this is seen in the *UGT1A1* gene. The *UGT1A1*28* (*NC_000002.12:g.233760235TA[8]*) and *UGT1A1*36* (*NC_000002.12:g.233760235TA[6]*) have different numbers of thymine adenine (TA) repeats in the TATA box region of the *UGT1A1* promoter regulating the expression of *UGT1A1* [7]. The effect of this is a decreased function of *UGT1A1*28* and an increased function of *UGT1A1*36*. Secondly, non-coding variants are suggested to alter RNA processing. For instance, rs12769205 (NC_000010.11:g.94775367A>C/G/T) located in the intron of *CYP2C19*, is one of the core variants of a loss-of-function allele, *CYP2C19*2*. This A>G alteration causes abnormal splicing, which is spliced mainly in the intron inclusion form rather than the normal form [8]. These examples warrant further investigations on how non-coding variants affect pharmacogenes function.

Notably, long-read sequencing not only detects the presence of variants, but also helps elucidate the mechanisms by which these variants influence enzyme activities. For instance, Ambrodji et al., sequenced the complete coding sequence (cDNA) of *DPYD* (4.4 kb) using Nanopore technology [9]. They found that 58% of reads in an individual carried the heterozygous c.1905+1G>A variant (rs3918290), a well-characterized risk variant, contained a 165 bp deletion in exon 14 [10]. In general, long-read sequencing is a useful tool to understand the mechanism of variants, especially genetic variants that cause splicing alterations.

Different types of long-read sequencing

In **Chapter 3 and 4**, we explored two strategies for implementing long-read sequencing: the capture-based approach and the amplicon-based approach, both using Pacific BioScience HiFi technology. The former improves sequencing efficiency compared to whole genome sequencing (WGS) by using capture probes to enrich targeted genes. With customized probes, captured genes can be adapted to research objectives, enabling the exploration of multiple genes or regions of interest. Yet, capture-based long-read sequencing panels remain costly due to the expenses for capture probes development and the larger size of the targeted sequence. Conversely, in **Chapter 4**, we developed a more specific amplicon-based approach for the *CYP3A* locus. This approach amplifies and enriches the target gene via PCR amplification. Multiple samples can be pooled in the same sequencing run by adding unique barcodes to amplified fragments, which substantially decreases the costs. Nonetheless, this method results in a smaller target, is relatively labor-intensive depending on the degree of PCR multiplexing and still costly if a large number of amplicons are needed. Both strategies offer flexibility in sequencing implementation, and the considerations for method selection should include objectives, budget, time and labor intensity.

Phasing

Haplotype phasing infers which variants are located on the same allele. It is crucial to construct haplotypes and link haplotypes to drug response. Phasing information is essential for evaluating the effect of novel impactful variants, especially in the intermediate metabolizer (IM). IM carries one normal function allele and one allele with decreased or completely lost function (Figure 7.2A). The impact of a heterozygous variant differs depending on whether it is located on the normal function allele or

on the loss-of-function allele. In **Chapter 3**, a potentially impactful *CYP2C19* variant NC_000010.11:g.94820597A>G (CADD score: 15.9) was detected in an intermediate metabolizer (*1/*2). This additional variant was assigned together with **2* based on haplotype phasing, resulting in a non-significant impact on CYP2C9 enzyme activity, as the **2* already has no residual activity. Should it have been located on the allele currently assigned a **1* call, this additional impactful variant will result in a decreased function allele. There are several methods for haplotype phasing. Firstly, computational phasing based on linkage disequilibrium is commonly used and correct on a population level. Population-based phasing algorithms determine the haplotype according to statistical linkage disequilibrium (LD) patterns. This approach depends on an assumption that the sequencing data are of the same or "similar" population to process the reliable LD information. But it could be problematic to process rare variants and novel variants which have not been observed previously [11]. Moreover, even for common variants, LD-based phasing is likely to be incorrect because phasing is processed based on a pre-defined LD threshold such as $r^2 > 0.8$ or $r^2 > 0.85$, which does not guarantee complete linkage. An r2 close to 1 indicates a strong linkage where the variants are very likely to be inherited together, while an r^2 closer to 0 suggests little or no association between them. The chosen r^2 threshold (e.g., 0.8 or 0.85) acts as a cutoff to determine which pairs of variants are likely to be inherited together. Higher thresholds (like 0.85) imply a stronger correlation, which can lead to more accurate phasing when variants are closely linked. However, even with high r^2, some variants that appear highly correlated in one population may not be as closely linked in another due to differences in recombination patterns or population history. For instance, *CYP2B6*4* (NC_000019.10:g.41009358A>G) variant and *CYP2B6*9* (NC_000019.10:g.41006936G>T) variant are in high LD, and are always assumed to be located on the same allele when both occur, resulting in a *6 annotation. However, in 1.5% of the patients carrying both variants, the two variants are on different alleles [12]. Even between variants assumed to have perfect LD ($r^2 = 1$), they do not always occur together. For example, the *DPYD HapB3* variant rs75017182 (NC_000001.11:g.97579893G>C) located in intron 10, which is the causal variant of *HapB3*, was regarded as having perfect linkage with rs56038477 (NC_000001.11:g.97573863C>T). Nonetheless, individuals with rs56038477 but not rs75017182 have been reported in the Children's Mercy Data Warehouse and the All of Us Research Program version 7 cohort [13], suggesting that testing rs56038477 only would generate false-positive genotype, resulting in underdose and a negative impact on the survival prospect. In their data, 0.223% of subjects with c.1236G>A lack c.1129-5923G>A. Notably, the data of both the mother and the child carrying c.1236G>A but lacking c.1129-5923G>A suggested that this observed haplotype is heritable rather than

a de novo or somatic variation. This case highlights the importance of more accurate phasing on a patient level.

In contrast to computational phasing, direct phasing resolves haplotype blocks with neighboring heterozygous variants gathered from sequencing. As a result, long-read sequencing is needed, especially in complex regions [14–17]. With read lengths reaching ~10 kb or longer, it is highly likely that sufficient heterozygous variants are included in a single read, and as a result, different variants can be phased together to their allele of origin without the need for computational approaches. In the 44 samples sequenced in **Chapter 3**, we achieved an overall on-target phasing of 62%. Notable, phasing performance can be improved by long-read sequencing compared to short-read sequencing (~20% phased on average) and varies between genes. This different phasing outcome between genes could be attributed to the occurrence of heterozygous variants as heterozygous variants are needed for the successful construction of haploblocks.

Moreover, phasing is of importance as it is more and more understood that enzyme function is affected by specific combinations on variants in one allele which interact together. The increased use of sequencing in PGx testing is expected to detect multiple variants, including impactful variants besides *-allele variants. If a patient is heterozygous for two deleterious variants, it is important to know if they are located on the same allele, resulting in one active and one inactive allele, or on opposing alleles, resulting in two non-active alleles (Figure 7.2B).

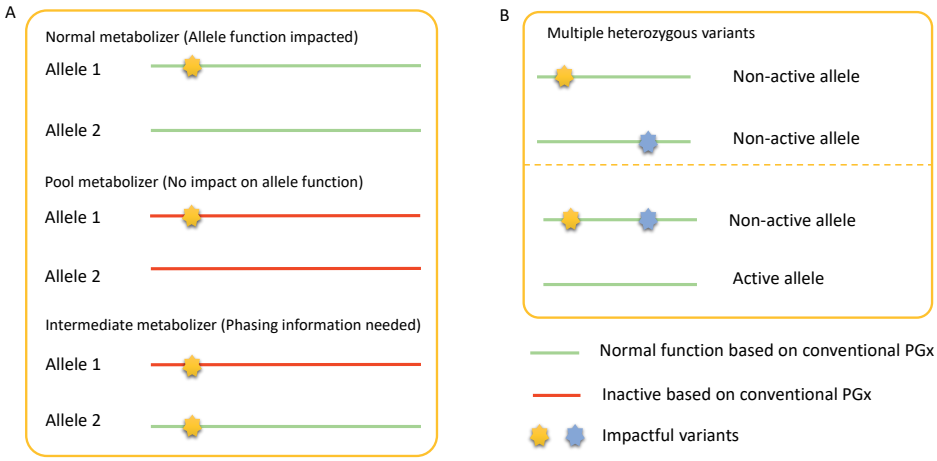


Figure 7.2: The importance of phasing information over predicted phenotypes.
(A) The effect of impactful variants on alleles with different function assigned by conventional PGx. **(B)** The effect of multiple heterozygous variants.

While direct phasing offers benefits for novel variants, it is not yet standard practice in pharmacogenomics. Computational phasing is still predominantly used by star(*)-allele tools. Most of these tools, including Stargazer [18], Aldy [19], StellarPGx [20] and PharmCAT [21], support a panel of pharmacogenes, while others are designed for a specific gene, such as Cyrius for the complex *CYP2D6* calling [22]. Depending on the input data, most of the tools are able to process BAM files, except PharmCAT. Moreover, there are more differences between these which impact how they are applied, including reference genomes, working environments, sequencing technologies and output information, which has been indicated by our study [23]. For instance, StellarPGx only supports short-read whole genome sequencing data. Of all the tools, Aldy is the only one which provides specific sequencing profiles, including long-read sequencing data [24]. Lastly, the type of output varies between tools. PharmCAT reports not only the diplotype but also the corresponding phenotype categorization based on the Clinical Pharmacogenetics Implementation Consortium (CPIC). Additionally, it shows the accompanying recommendations from the CPIC guidelines, which are useful for clinical use. In contrast, other tools such as Aldy does not include clinical interpretation but it reports core *-alleles and minor *-alleles as well as all variants processed, which is valuable for research (Table 7.1).

Table 7.1: Overview of star(*)-allele tools and their specifications

		Stargazer	Aldy	StellarPGx	PharmCat	Cyrius
Input data	BAM	√	√	√		√
	CRAM	√	√	√		
	VCF	√	√		√	
Target genes	Multiple	√	√	√	√	
	Single					*CYP2D6*
Software requirement	Linux OS	√	√	Singularity		√
	Mac OS	√	√	Docker		Python
	Windows	√			Java	
Sequencing data supported	NGS (WGS)	√	√	√	√	√
	LRS		√			
	SNP array	√				
Reference database	PharmVar	√	√	√		
	PharmGKB		√		√	
	CPIC and/or DPWG				√	

Abbreviations: NGS, next-generation sequencing; LRS, long-read sequencing; CPIC, Clinical Pharmacogenetics Implementation Consortium; DPWG, Dutch Pharmacogenetics Working Group.

Several studies have evaluated the performance of these tools. Most discrepancies were observed in the *CYP2D6* gene, in which the specifically designed Cyrius showed the most robust performance [25]. Importantly, sequencing depth plays a crucial role in the accuracy of *-allele calling. Lower depths, particularly < 5X lead to reduced accuracy. Whereas, increased depth higher than 20X results in only minor improvements in most genes [25]. The performance of *-allele calling tools is highly accurate in general if users are able to select proper *-allele calling tools according to the sequencing methods, usage and outcome needs.

High-throughput variants interpretation

With the long-read sequencing approaches described in **Chapters 3 and 4**, a great number of variants were found across multiple pharmacogenes. Nonetheless, only a small number of pharmacogenetic variants (about 3%) have been assigned to star(*)-allele nomenclature (about 3%) in our data. The remaining 97% of variants detected is not included in the *-alleles. Conventionally, one of the gold standard strategies to address the functionality of variants is experimental functional assays, conducted by introducing novel variants into cell lines to quantitatively assess the gene expression. Subsequently, the clinical consequences can be evaluated by clinical studies which connect the presence of variants to the corresponding phenotypes *in vivo*. However, these approaches are not always feasible given the large number of variants detected via high-throughput sequencing technologies and the relatively low allele frequency of many variants in pharmacogenes [26–28]. This is a limitation for *in vivo* studies in particular, given the extremely high number of patients needed to access the clinical impact of rare variants.

Currently, computational tools are emerging as a promising solution for the functional prediction of a large number of genetic variants. This approach is also less costly and time-consuming compared to experimental assays. To date, multiple algorithms based on diverse parameters, such as sequence conservation and structure features, have been developed to predict variant effect [27]. While some tools (e.g., CADD score) are designed for general types of variants, most of the tools were trained to predict a specific type of variants. For instance, SIFT and PolyPhen2 focus on missense variants and MMsplice is specific for splice variants [29]. Those tools were developed according to different features of different functional regions in the genome. For instance, for variants in the coding sequences, conservation and structure stability are of great importance for function prediction. By contrast, variants affecting transcription factor binding and polymerase loading are more relevant for predicting regulatory effects [27].

The performance of 18 different functionality prediction tools, each based on diverse features, has been evaluated [30]. Notably, a large difference in the predictive specificity and sensitivity between tools was observed when testing variants, indicating the importance of tool selection and careful evaluation of the prediction results. Moreover, it is common for these algorithms to perform poorly when tested on pharmacogenetic variants than variants related to the disease, as these models were trained based on disease traits, which have a great discrepancy with drug responses. In a study by Zhou et al., some algorithms performed less accurately on pharmacogenetic variants compared to their training data, as they were primarily designed to identify the pathogenicity of variants (disease related) rather than functional consequences. Moreover, many pharmacogenes are poorly conserved, which limits the effectiveness of algorithms based on evolutionary constraints [30]. In **Chapter 3**, the CADD score was used to predict the impact of variants found by our long-read panel as it was one of the tools with the best performance among 18 models assessed by Zhou et al. In our analysis of selected known function variants in pharmacogenes, the CADD score presented a sensitivity of 77% and specificity of 59%.

Additionally, the performance of computational tools in non-coding variants is questionable, probably because of the complexity of their effects [29]. In our data (**Chapter 3**), CADD score failed to identify rs12248560 (NC_000010.11:g.94761900C>T), a well-known variant characterizing the increased function allele *CYP2C19*17*, as a high-impact variant. This issue also occurred with other non-coding variants associated with *-alleles with known non-normal activity such as *CYP2C19*2* (rs12769205) and *UGT1A1*28* (rs3064744). These findings suggest that the *in silico* identification of impactful variants in non-coding regions remains challenging. To address this problem, several *in silico* tools were developed specifically for variants located in non-coding regions, which account for > 99% of the human genome [4, 29, 31]. Examples include DeepSEA [32] for regulatory variants and ImiRP [33] for variants in untranslated regions. However, the impact of these abundant variants is substantially under-studied compared to variants in protein-coding regions such as missense variants and frameshift variants [27].

Besides general and variant type-specific computational algorithms, a few computational methods were developed to predict the variant effect in specific genes. These tools are often assumed to have a better performance than the non-specific tools [29]. For instance, DPYD-Varifier which was applied in **Chapter 6**, has a prediction accuracy of over 85% [34]. However, the scope of such gene-specific tools should be considered. In **Chapter 6**, the DPYD-Varifier was used to predict the *DPYD* variants in the coding regions only, as the tool was trained based on the *in vitro* DPD activity of

156 missense variants. Conversely, the remaining splice site variants were interpreted using the MMsplice [35]. The features processed by computational tools for missense variants and splice site variants are different, which determines the application scope of different algorithms.

In addition to the computational prediction, the *in vitro* experimental assays and clinical consequences are crucial to further confirm the effects of the variants predicted to be deleterious. In **Chapter 6**, the function of nine *DPYD* missense variants was estimated by the *in vitro* expression system and the outcome was highly consistent with the DPYD-Varifier prediction. Only one variant, c.2194G>T_p.Val732Phe, was predicted to be deleterious via DPYD-Varifier but *in vitro* assessment reported it as a neutral variant. While accurate, such an *in vitro* system is not feasible for studying a larger number of variants. Recently, deep mutational scanning [36] and massively parallel reporter assays [37] have been developed, allowing the simultaneous assessment of a large library of variants to overcome the low-throughput of conventional experimental assays. These assays have been applied to several pharmacogenes such as *TPMT* [38] and *CYP2C9* [39], and are regarded as powerful methods to assess variant effects *in vitro*. Nonetheless, the clinical associations are difficult to determine, given the low occurrence of rare variants or the marginal effect size. Most studies are underpowered or irreproducible, resulting in low levels of clinical evidence, and therefore, they do not meet the prerequisites of clinical utilization or guidelines. The case-control approach has been applied in multiple studies, suggesting it as a solution for addressing the contribution of rare variants [40–42]. However, those studies typically enrolled a large number of cases and controls (thousands patients enrolled). As a result, most of studies heavily relied on data derived from national biobanks or large-scaled public databases. The detailed drug prescribing information and clinical outcomes needed for PGx studies are not widely available in those sources. Prospective data collection are more likely to get high quality data, for instance, the cohort in **Chapter 6** (the ALPE-DPD study) formed a comprehensive and well-structured dataset. However, the sample size of the ALPE study was calculated based on the primary objective of identifying the influence of four deleterious *DPYD* variants, which might be insufficient to investigate rare variants. In **Chapter 6**, we therefore applied matched-pair methods to compromise the influence of several related factors on fluoropyrimidine-induced severe toxicity. After excluding the four main variants which are generally tested for, there were seven additional *DPYD* variants predicted to be deleterious based on *in vitro* expression system and MMSplice. However, these seven deleterious *DPYD* variants could not be linked to severe toxicity. Meanwhile, prospective clinical trials can be time-consuming and costly given the large

sample size required to detect variants with low frequency. In contrast, retrospective studies are less labor-intensive and more cost-effective, but they might be limited by the completeness and quality of data.

Intricate regulation of pharmacogenes expression

In addition to variants within a gene that directly impact its function, gene function is also determined by its expression level. Furthermore, pharmacogenes, especially *CYP* genes, are characterized by dramatic inter-individual variability in expression according to the GTEx data (Figure 7.3). Theoretically, gene expression is intricately regulated by transcription factors (TFs). Functional variants in TFs can alter their binding affinity to promoters of downstream genes or result in up- or downregulation of TF expression. Yet, previous studies show contradictory results regarding the effects of some variants on the enzyme activity and substrate metabolism due to insufficient *in vivo* evidence. For instance, a variant might have no influence on the substrate metabolism *in vivo*, despite altering the expression of the transcription factor and the enzyme activity *in vitro* [43, 44], which highlights the importance of *in vivo* evaluation. Similarly, in **Chapter 5**, although the effect of the variants in the transcription factors was validated *in vitro*, these effects are potentially compromised by confounding factors *in vivo* as we were unable to replicate the effect *in vivo*.

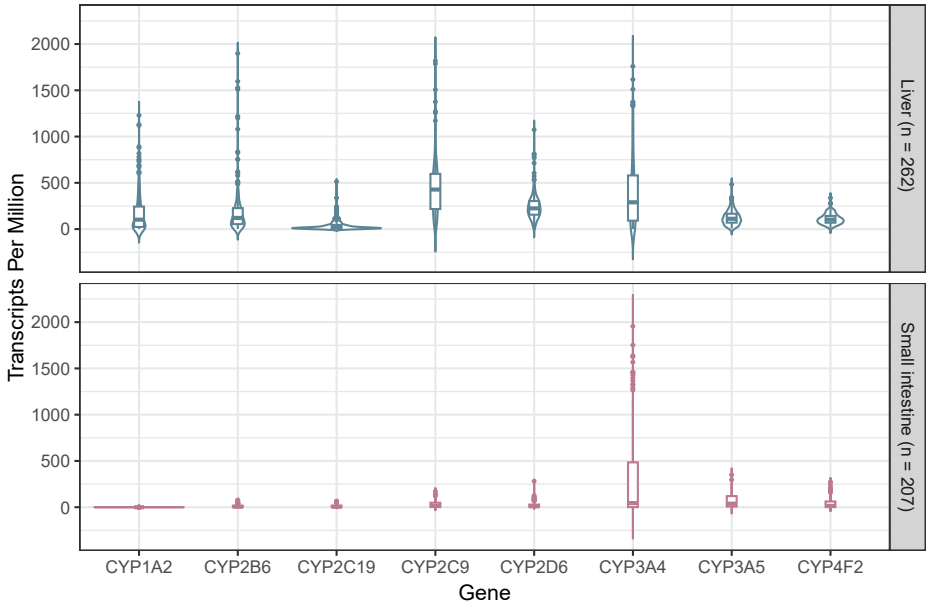


Figure 7.3: Expression level of CYP enzymes in liver and small intestine.

Apart from transcription regulation, post-transcriptional and post-translational modification play a role in gene expression. Post-transcriptional regulation mainly refers to the microRNAs (miRNA). For example, miR-30c-1-3p, has been reported to alter the PXR expression by targeting the 3'-untranslated region (UTR) of the PXR transcripts, thereby, it decreases the mRNA level of PXR and CYP3A4 [45]. Numerous miRNAs have been implicated in the regulation of various pharmacogenes, and their regulatory effect on the drug metabolizing enzymes is also influenced by genetic variants, the presence of pseudogenes, and RNA editing [46]. Post-translational modifications, such as phosphorylation of *CYP2D6* and ubiquitination of *CYP3A4*, are diverse and have not been widely studied in PGx [47]. Nonetheless, they hold promise for explaining the missing heritability of drug responses.

Beyond single drug-gene interaction

Given the complexity of the variant effect on the variability of drug response, it is apparent that the individual alterations in drug response cannot be fully explained by single variant interactions. This highlights the need to investigate the integrated effects of multiple genes and variants [48–51]. Some genes that are involved in the metabolic pathway can have their potential associations identified by analyzing functional variants in candidate genes. In contrast, for genes with an indirect relevance to drug response, in which their role is not yet clearly addressed, genome-wide association studies (GWASs) offer an option to screen the candidate genes and identify relevant loci/genes. The use of GWAS improves the collective knowledge of the effect of genetics on drug response, by confirming findings from candidate gene studies and identifying novel drug-gene or drug-variant associations [52–54]. Despite those findings, PGx-related GWAS has differences compared to disease GWAS (Table 7.2) and encounters several challenges. Firstly, most GWASs use a widely accepted *p*-value threshold of 5×10^{-8} [55], suggesting that only common variants with large effect size are able to be identified as significant. However, variants with small-effect and moderate-effect also play a role in drug response. Muhammad et al. indicated that these variants explained more than half of the narrow-sense heritability in the given pharmacodynamic and pharmacokinetic phenotypes [50]. Moreover, most GWASs in PGx did not account for rare variants or even low-frequency variants (MAF ≤ 5%) due to the quality control or limitations in analytical power [56]. Nonetheless, the number of low allele frequency variants presents an enrichment in drug response signals rather than common diseases and other complex traits [57], highlighting the contribution of rare variants in drug response. Secondly, collecting a cohort with thorough drug exposure and phenotyping

data is challenging, even if a small sample size would be sufficient due to the large effect size of variants/genes involved in the drug metabolism pathway or drug target directly. Lastly, the increasing use of GWAS encourages the requirements of variant interpretation as aforementioned, and translating functional variants into clinically actionable recommendations is challenging, especially for the rare pharmacogenetic variants [58].

Table 7.2: Characteristics of pharmacogenomics (PGx) GWAS and disease GWAS

	PGx GWAS	Disease GWAS
Objective	Identify genetic variants that related to inter-individual variability of drug-response	Identify genetic variants that associated with the risk of disease occurrence
Phenotype definition	Drug response, including efficacy, adverse events, pharmacokinetics parameters et al.	A disease diagnosis, severity
Study design	Case-control design and cohort design (following all treated patients)	Case-control design (Individuals with disease VS without disease)
Sample size	Smaller samples size is possible depending on the effect size of variants	Larger population size required
Study cohort	Challenging in finding well-characterized cohorts	Large-scaled population and biobank data
Polygenic	Both monogenic and polygenic	Highly polygenic
Important confounding factors	Patients adherence	Environmental factors

To further gather the effect of multiple genetic variants on phenotypes, polygenic risk scores (PRS), originally derived from GWAS to predict the risk of human diseases or complex traits are considered. In PGx, the PRS could also be useful, when applied with drug response as the phenotype. A PRS aggregates multiple single nucleotide variants via weighting their effect size from an independent discovery GWAS. This approach provides a comprehensive measure of the total genetic contribution on a particular phenotype, including not only variants with large effects but also those with moderate or minor effects, which cannot reach the GWAS significant threshold [59, 60]. Compared to *-allele nomenclature, which is able to contain multiple variants, PRS includes variants from different genes. The applications of PRSs in PGx have been reviewed [61–63], showing their promising utility. Nonetheless, over 80% of currently published studies that use PRS are based on disease GWAS, which use SNPs that are not directly implicated in drug response [60, 61]. For instance, in a large-scaled GWAS involving over 11,000 coronary artery disease cases, PRS was significantly associated

with the incidence of major adverse cardiovascular events (MACE) and the risk reduction of MACE after alirocumab treatment [64]. Moreover, PRSs derived from disease GWAS do not fully cover the heritability of drug response, despite the broader resources and typically larger sample sizes [60, 65]. So using data from the base cohort of disease GWASs (the cohort used to generate summary statistics to obtain variant effect sizes and *p*-values) results in about 20% failure prediction, which means no significant associations between PRS and drug response. In contrast, the failure rate decreased to 6% with data from PGx GWAS [60]. Recently, Zhai et al. processed a PRS-PGx method that improves the prediction accuracy of PRS for drug response but it has not been widely used [65]. Briefly, their model was developed by giving not only the prognostic effect but also the predictive effects of variants to make the PRS shift to the PGx PRS approach.

Future prospectives

With advancements in genome sequencing technologies and high-throughput data processing methods, PGx has a broad implamentation in predicting drug response. There are promising directions for future studies.

Firstly, predicting drug response is a complex system involving multiple intermediate layers as Figure 7.4 shows [66]. Advances in biotechnology and bioinformatics are driving PGx studies to integrate multidimensional, large-scale omics data, including the genome, transcriptome, proteome, and epigenome. By leveraging multi-omics data, we are able to have comprehensive insights into the missing heritability and molecular mechanisms underlying complex drug response. For instance, epigenetic mechanisms modifying the gene expression without altering the DNA sequence, can help to clarify the missing heritability in PGx [67–71]. Multiple studies have indicated that the hypermethylation of the ADME gene promoters suppresses their expression, while the hypomethylation has the opposite effect [72]. However, to date, the contribution of epigenetics on drug response has not been well-characterized due to the limited application of epigenetic testing and the complexity of epigenetic principles [73]. For instance, liver tissue might be needed for liver-specific expression assessment. Moreover, epigenetic modifications are diverse, including DNA methylation, histone modification, non-coding RNA-mediated regulation as well as the chromatin remodeling [73]. Notably, the epigenetic-mediated drug response is determined not only by basal epigenetic states but also by environment-induced modification, specifically, drug administration [74–76]. Some common drugs such as opioids and valproic could induce epigenetic alterations, resulting in wanted and unwanted drug effects [75].

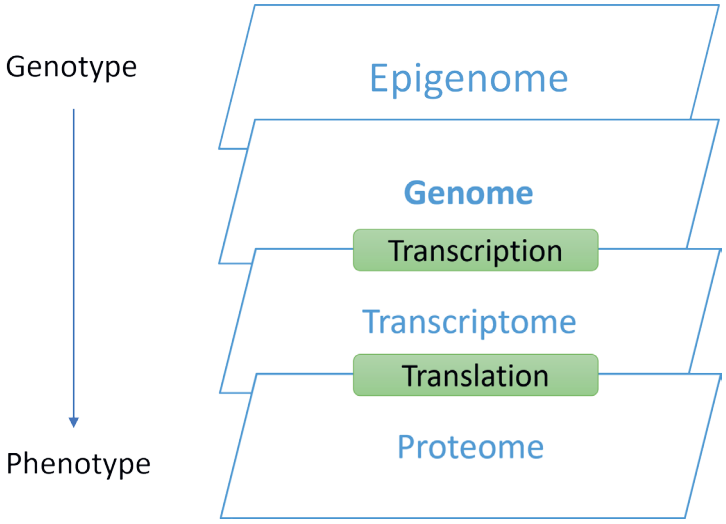


Figure 7.4: Layers for biology mechanisms of drug response.

About 5% of all drugs are potentially related to human histone deacetylase according to *in vitro* experiments. Further investigations via fundamental or clinical experiments are expected to address drug-induced epigenetics changes and how epigenetic impacts drug response.

Sample size determines the statistical power in PGx studies, which is of great importance to draw solid conclusions. Additionally, extremely large cohorts may be necessary to evaluate the contribution of rare variants, which can be costly. Zuk et al.'s study indicated that > 25,000 cases, should be involved in the discovery sets and the substantial replication set to have a well-powered rare variant association study [77]. Meanwhile, the increasing availability of population-based large-scale biobanks offers a more feasible and cost-effective approach by providing cohorts containing sufficient participants [66]. The use of biobank data has huge advantages, and it can be further strengthened by integrating phenotype information, for instance, from electronic health care records (EHRs). Tase et al.'s study identified novel loss-of-function variants and missense variants in 64 very important pharmacogenes (VIP) with EHRs-coupled biobank whole-genome sequencing data with over 2,000 participants [78]. In total, 41 loss-of-function and 567 missense variants were detected, and the frequency of most detected variants was below 0.05%. This study replicated several previously reported variant-drug associations and identified novel interactional signals, for instance, rs145259190 (NC_000006.12:g.160262532C>T), located in the promoter

region of *SLC22A2*, was related to metformin toxicities [78]. However, studies based on biobank data could also be challenging. Firstly, the quality of the corresponding clinical data varies greatly and often contains unstructured, sparse and incomplete records, especially for some clinical outcomes that require manual evaluation and annotation. Secondly, new findings identified with biobank data are insufficient to prove causality, and further validation is needed before they could be deemed clinically actionable. Therefore, conventional approaches, such as randomized clinical trials (RCTs), are still considered to be a higher level of evidence over data-driven approaches.

The growing availability of large-scale biobanks and the multi-omics data highlight the potential benefits of artificial intelligence (AI) modeling in PGx studies [79–81]. Machine-learning (ML), as a subset of AI, is rapidly developing in medicine [82]. For instance, a study by Gottlieb et al., explained an additional 8–12% of the variation in warfarin pharmacokinetics (PK) by integrating the expression data of 116 genes related to warfarin PK and PD pathways [83]. Multiple ML-based algorithms have been developed to predict variant effects and an ensemble ML model has shown superior accuracy in predicting variants in ADME genes [30]. Notably, among techniques within ML, the neural network is more specifically implemented in managing genotype-phenotype analysis and mutation-gene-drug relations [84]. The prediction of CYP2D6-mediated tamoxifen metabolism can be improved by using a neural network based continuous scale phenotype instead of the conventional categorical phenotype [85].

Solid evidence indicates that individuals from different ethnic groups exhibit varied responses to specific agents [86]. One explanation is the different frequencies of impact variants between ethnic groups, including common variants and rare variants [87, 88]. Despite these differences, PGx research and implementation have significant biases between populations, of which the outcomes are predominately focused on European ancestry [86, 89, 90]. These disparities in PGx testing prevalence and quality underscore broader healthcare inequalities. Therefore, conducting studies with more diverse populations is crucial to reduce bias and identify novel drug-gene associations by linking variations in drug response to genetic diversity.

Current PGx studies investigated more factors implicated in pharmacokinetics (PK) than pharmacodynamics (PD). Moreover, current PGx guidelines primarily focus on PK-related dose adjustments, which address only one part of the overall picture. In contrast, pharmacodynamics, which examines interactions with the drug targets or the biological pathway, plays an important role in drug response, particularly drug efficacy. In contrast to PK measurement, examination of PD has multiple difficulties.

Firstly, identifying the treatment benefit can be difficult due to issues like the appropriate endpoint, accounting for the placebo effect and ensuring patient adherence [91]. Secondly, besides the interaction with dosing [92, 93], drug efficacy is intensively related to more specific factors, such as the drug target or the corresponding biological pathway that alters the sensitivity to the drug. To study how genetic variants affect PD, disease-specific, drug-specific and tissue-specific considerations are to be taken into account. Regarding drug targets and pathways, anticancer agents represent a special case in PGx, as responses to anticancer agents involves two genomes: the germline genome and the somatic genome. Notably, in most clinical settings, somatic variants provide identification of the malignancy types that are likely to respond to the anticancer agents. Therefore, the corresponding agents are prescribed to patients carrying acquired variants, and the upfront genetic testing is performed in tumor tissue only [94]. As a result, germline testing remains underutilized in cancer treatment. Notably, with GWAS, only 15% of drugs or drug classes were identified to have robust genetic associations for efficacy [91]. But this observation could be underestimated due to the difficulties mentioned above, which implicates a higher demand for study design in future PGx research.

Additionally, the mismatch between genotype and phenotype has been widely observed, which is referred to as phenoconversion. In some studies, phenoconversion has been regarded as an explanation for the lack of association between genotype and phenotype in CYP enzymes [95]. The reasons for the discordance are diverse, for instance, drug-drug interactions (DDI), special physiological and pathophysiological situations such as inflammation and other factors such as smoking [96]. Klomp et al.'s study found that higher C-reaction protein (CRP) level (an inflammatory biomarker) is related to lower CYP2C19 enzyme activity compared to the genetics-predicted phenotype, regardless of the genotype. Meanwhile, CYP2C19 activity showed a similar pattern of variation as the CRP fluctuation [97].

There are multiple methods for phenotype observation. Phenotypes like enzyme activity are not always possible to assess directly. Consequently, surrogate biomarkers have been widely used for the sake of feasibility and cost-effectiveness, with liquid biopsy showing particular advantages. Blood contains liver-released extracellular vesicles (EVs), which contain functional protein, mRNA and metabolites. Liquid biopsy isolates these EVs from blood to allow biomarker detection in a less invasive way than a tissue biopsy. Herein, testing the components in EVs could provide liver-specific enzyme activity data and gene expression data. Multiple CYP enzymes (both mRNAs and proteins) have been detected in plasma-derived exosomes [98]. Rowland et al. indicated that

CYP3A4 enzyme activity could be surrogated by CYP3A4 expression in exosome as the midazolam clearance was significantly associated with exosomal CYP3A4 protein levels (R^2 = 0.91), and CYP3A4 mRNA levels (R^2 = 0.79) [99]. These studies indicate a relatively high concordance between enzyme activities and surrogated biomarkers derived from liquid biopsy, suggesting a promising implementation in PGx.

The human gut microbiome is one of the missing fields that might have more complex interactions with drug metabolism pathways than previously addressed [100]. Increasing evidence shows that gut microbiome significantly influences pharmacokinetics by altering all four components of ADME [101, 102]. For instance, gut bacteria process β-glucuronidases, which are involved in the biotransformation of the inactive metabolite of irinotecan to an active metabolite (SN-38). This active component is the cause of diarrhea, occurring in up to 40% of patients [103]. Besides, gut microbiome is related to pharmacodynamics by interacting with metabolism and immunity, as seen in the response to immune checkpoint inhibitors [104, 105]. Notably, the role of the gut in drug response varies significantly due to microbiome composition, variations of the host genome, and environmental factors [106], and integrating gut into current PGx knowledge might improve the explanation of missing heritability.

As all aspects aforementioned, predicting drug response with PGx is complex. It requires a comprehensive understanding of genome make-up of genes involved directly, networking of genes has indirect effects, and taking non-genetic factors into account. To investigate the complex interactions between the human genome and drug response, new technologies and large-scale data are necessary but lead to challenge of data processing and analysis. Nevertheless, advances in PGx and personalized medicine hold immense potential to revolutionize healthcare, enabling safer and more effective treatments tailored to each individual's unique genetic profile.

References

1. Swen JJ, van der Wouden CH, Manson LE, et al. A 12-gene pharmacogenetic panel to prevent adverse drug reactions: an open-label, multicentre, controlled, cluster-randomised crossover implementation study. Lancet. 2023;401(10374):347-356.
2. Lauschke VM, Milani L, Ingelman-Sundberg M. Pharmacogenomic Biomarkers for Improved Drug Therapy-Recent Progress and Future Developments. AAPS J. 2017;20(1):4.
3. van der Lee M, Rowell WJ, Menafra R, Guchelaar HJ, Swen JJ, Anvar SY. Application of long-read sequencing to elucidate complex pharmacogenomic regions: a proof of principle. Pharmacogenomics J. 2022;22(1):75-81.
4. Gloss BS, Dinger ME. Realizing the significance of noncoding functionality in clinical genomics. Exp Mol Med. 2018;50(8):1-8.
5. Gaedigk A, Casey ST, Whirl-Carrillo M, Miller NA, Klein TE. Pharmacogene Variation Consortium: A Global Resource and Repository for Pharmacogene Variation. Clin Pharmacol Ther. 2021;110(3):542-545.
6. Wu K, Bu F, Wu Y, et al. Exploring noncoding variants in genetic diseases: from detection to functional insights. J Genet Genomics. 2024;51(2):111-132.
7. Peterkin VC, Bauman JN, Goosen TC, et al. Limited influence of UGT1A1*28 and no effect of UGT2B7*2 polymorphisms on UGT1A1 or UGT2B7 activities and protein expression in human liver microsomes. Br J Clin Pharmacol. 2007;64(4):458-468.
8. Chiang HL, Wu JY, Chen YT. Identification of functional single nucleotide polymorphisms in the branchpoint site. Hum Genomics. 2017;11(1):27.
9. Ambrodji A, Sadlon A, Amstutz U, et al. Approach for Phased Sequence-Based Genotyping of the Critical Pharmacogene Dihydropyrimidine Dehydrogenase (DPYD). Int J Mol Sci. 2024; 25(14).
10. van Kuilenburg AB, Meijer J, Mul AN, et al. Intragenic deletions and a deep intronic mutation affecting pre-mRNA splicing in the dihydropyrimidine dehydrogenase gene as novel mechanisms causing 5-fluorouracil toxicity. Hum Genet. 2010;128(5):529-538.
11. Tewhey R, Bansal V, Torkamani A, Topol EJ, Schork NJ. The importance of phase information for human genomics. Nat Rev Genet. 2011;12(3):215-223.
12. van der Lee M, Allard WG, Bollen S, et al. Repurposing of Diagnostic Whole Exome Sequencing Data of 1,583 Individuals for Clinical Pharmacogenetics. Clin Pharmacol Ther. 2020;107(3):617-627.
13. Turner AJ, Haidar CE, Yang W, et al. Updated DPYD HapB3 haplotype structure and implications for pharmacogenomic testing. Clin Transl Sci. 2024;17(1):e13699.
14. Wenger AM, Peluso P, Rowell WJ, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol. 2019;37(10): 1155-1162.
15. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. Nature Reviews Genetics. 2018;19(6):329-346.
16. Cretu Stancu M, van Roosmalen MJ, Renkens I, et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. Nat Commun. 2017;8(1):1326.

17. Scott ER, Yang Y, Botton MR, et al. Long-read HiFi sequencing of NUDT15: Phased full-gene haplotyping and pharmacogenomic allele discovery. Hum Mutat. 2022;43(11):1557-1566.
18. Lee SB, Wheeler MM, Patterson K, et al. Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. Genet Med. 2019;21(2):361-372.
19. Hari A, Zhou Q, Gonzaludo N, et al. An efficient genotyper and star-allele caller for pharmacogenomics. Genome Res. 2023;33(1):61-70.
20. Twesigomwe D, Drögemöller BI, Wright GEB, et al. StellarPGx: A Nextflow Pipeline for Calling Star Alleles in Cytochrome P450 Genes. Clin Pharmacol Ther. 2021;110(3):741-749.
21. Sangkuhl K, Whirl-Carrillo M, Whaley RM, et al. Pharmacogenomics Clinical Annotation Tool (PharmCAT). Clin Pharmacol Ther. 2020;107(1):203-210.
22. Chen X, Shen F, Gonzaludo N, et al. Cyrius: accurate CYP2D6 genotyping using whole-genome sequencing data. Pharmacogenomics J. 2021;21(2):251-261.
23. Graansma LJ, Zhai Q, Busscher L, et al. From gene to dose: Long-read sequencing and *-allele tools to refine phenotype predictions of CYP2C19. Front Pharmacol. 2023;14:1076574.
24. Numanagic I, Malikic S, Ford M, et al. Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. Nat Commun. 2018;9(1):828.
25. Halman A, Lunke S, Sadedin S, Moore C, Conyers R. Benchmarking pharmacogenomics genotyping tools: Performance analysis on short-read sequencing samples and depth-dependent evaluation. Clin Transl Sci. 2024;17(8):e13911.
26. Ingelman-Sundberg M, Mkrtchian S, Zhou Y, Lauschke VM. Integrating rare genetic variants into pharmacogenetic drug response predictions. Hum Genomics. 2018;12(1):26.
27. Zhou Y, Fujikura K, Mkrtchian S, Lauschke VM. Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data. Front Pharmacol. 2018;9:1437.
28. Kozyra M, Ingelman-Sundberg M, Lauschke VM. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. Genet Med. 2017;19(1):20-29.
29. Zhou Y, Lauschke VM. Computational Tools to Assess the Functional Consequences of Rare and Noncoding Pharmacogenetic Variability. Clin Pharmacol Ther. 2021;110(3):626-636.
30. Zhou Y, Mkrtchian S, Kumondai M, Hiratsuka M, Lauschke VM. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. Pharmacogenomics J. 2019;19(2):115-126.
31. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nat Rev Genet. 2015;16(4):197-212.
32. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods. 2015;12(10):931-934.
33. Ryan BC, Werner TS, Howard PL, Chow RL. ImiRP: a computational approach to microRNA target site mutation. BMC Bioinformatics. 2016;17:190.
34. Shrestha S, Zhang C, Jerde CR, et al. Gene-Specific Variant Classifier (DPYD-Varifier) to Identify Deleterious Alleles of Dihydropyrimidine Dehydrogenase. Clin Pharmacol Ther. 2018; 104(4):709-718.
35. Cheng J, Nguyen TYD, Cygan KJ, et al. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. Genome Biol. 2019;20(1):48.

36. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014;11(8):801-807.
37. van Arensbergen J, Pagie L, FitzPatrick VD, et al. High-throughput identification of human SNPs affecting regulatory element activity. Nat Genet. 2019;51(7):1160-1169.
38. Chiasson M, Dunham MJ, Rettie AE, Fowler DM. Applying Multiplex Assays to Understand Variation in Pharmacogenes. Clin Pharmacol Ther. 2019;106(2):290-294.
39. Amorosi CJ, Chiasson MA, McDonald MG, et al. Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. Am J Hum Genet. 2021;108(9):1735-1751.
40. Choi J, Kim S, Kim J, et al. A whole-genome reference panel of 14,393 individuals for East Asian populations accelerates discovery of rare functional variants. Sci Adv. 2023;9(32):eadg6319.
41. Sadeghi-Alavijeh O, Chan MMY, Moochhala SH, Howles S, Gale DP, Böckenhauer D. Rare variants in the sodium-dependent phosphate transporter gene SLC34A3 explain missing heritability of urinary stone disease. Kidney Int. 2023;104(5):975-984.
42. Glessner JT, Ningappa MB, Ngo KA, et al. Biliary atresia is associated with polygenic susceptibility in ciliogenesis and planar polarity effector genes. J Hepatol. 2023;79(6):1385-1395.
43. Moes DJ, Press RR, den Hartigh J, van der Straaten T, de Fijter JW, Guchelaar HJ. Population pharmacokinetics and pharmacogenetics of everolimus in renal transplant patients. Clin Pharmacokinet. 2012;51(7):467-480.
44. Du QQ, Wang ZJ, He L, Jiang XH, Wang L. PXR polymorphisms and their impact on pharmacokinetics/pharmacodynamics of repaglinide in healthy Chinese volunteers. Eur J Clin Pharmacol. 2013;69(11):1917-1925.
45. Vachirayonstien T, Yan B. MicroRNA-30c-1-3p is a silencer of the pregnane X receptor by targeting the 3'-untranslated region and alters the expression of its target gene cytochrome P450 3A4. Biochim Biophys Acta. 2016;1859(9):1238-1244.
46. Nakano M, Nakajima M. Current knowledge of microRNA-mediated regulation of drug metabolism in humans. Expert Opin Drug Metab Toxicol. 2018;14(5):493-504.
47. He ZX, Chen XW, Zhou ZW, Zhou SF. Impact of physiological, pathological and environmental factors on the expression and activity of human cytochrome P450 2D6 and implications in precision medicine. Drug Metab Rev. 2015;47(4):470-519.
48. Roden DM, McLeod HL, Relling MV, et al. Pharmacogenomics. Lancet. 2019;394(10197):521-532.
49. Crouch DJM, Bodmer WF. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. Proc Natl Acad Sci U S A. 2020;117(32):18924-18933.
50. Muhammad A, Aka IT, Birdwell KA, et al. Genome-Wide Approach to Measure Variant-Based Heritability of Drug Outcome Phenotypes. Clin Pharmacol Ther. 2021;110(3):714-722.
51. Damon C, Luck M, Toullec L, et al. Predictive Modeling of Tacrolimus Dose Requirement Based on High-Throughput Genetic Screening. Am J Transplant. 2017;17(4):1008-1019.
52. McInnes G, Yee SW, Pershad Y, Altman RB. Genomewide Association Studies in Pharmacogenomics. Clin Pharmacol Ther. 2021;110(3):637-648.

53. Low SK, Takahashi A, Mushiroda T, Kubo M. Genome-wide association study: a useful tool to identify common genetic variants associated with drug toxicity and efficacy in cancer pharmacogenomics. Clin Cancer Res. 2014;20(10):2541-2552.
54. Yee SW, Momozawa Y, Kamatani Y, et al. Genomewide Association Studies in Pharmacogenomics: Meeting Report of the NIH Pharmacogenomics Research Network-RIKEN (PGRN-RIKEN) Collaboration. Clin Pharmacol Ther. 2016;100(5):423-426.
55. Jannot AS, Ehret G, Perneger T. P < 5 x 10(-8) has emerged as a standard of statistical significance for genome-wide association studies. J Clin Epidemiol. 2015;68(4):460-465.
56. Mn W, L J, Jw H, et al. Use of SNP chips to detect rare pathogenic variants: retrospective, population based diagnostic evaluation. Bmj. 2021;372:n214.
57. Zhou K, Pearson ER. Insights from genome-wide association studies of drug response. Annu Rev Pharmacol Toxicol. 2013;53:299-310.
58. Zhou Y, Tremmel R, Schaeffeler E, Schwab M, Lauschke VM. Challenges and opportunities associated with rare-variant pharmacogenomics. Trends Pharmacol Sci. 2022;43(10):852-865.
59. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020;15(9):2759-2772.
60. Zhai S, Mehrotra DV, Shen J. Applying polygenic risk score methods to pharmacogenomics GWAS: challenges and opportunities. Brief Bioinform. 2023;25(1).
61. Siemens A, Anderson SJ, Rassekh SR, Ross CJD, Carleton BC. A Systematic Review of Polygenic Models for Predicting Drug Outcomes. J Pers Med. 2022;12(9).
62. Johnson D, Wilke MAP, Lyle SM, et al. A Systematic Review and Analysis of the Use of Polygenic Scores in Pharmacogenomics. Clin Pharmacol Ther. 2022;111(4):919-930.
63. Cross B, Turner R, Pirmohamed M. Polygenic risk scores: An overview from bench to bedside for personalised medicine. Front Genet. 2022;13:1000667.
64. Damask A, Steg PG, Schwartz GG, et al. Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial. Circulation. 2020;141(8):624-636.
65. Zhai S, Zhang H, Mehrotra DV, Shen J. Pharmacogenomics polygenic risk score for drug response prediction using PRS-PGx methods. Nat Commun. 2022;13(1):5278.
66. Auwerx C, Sadler MC, Reymond A, Kutalik Z. From pharmacogenetics to pharmaco-omics: Milestones and future directions. HGG Adv. 2022;3(2):100100.
67. Singh S, Kumar P, Kumar JP, et al. Genetic and Epigenetic Basis of Drug-Induced Liver Injury. Semin Liver Dis. 2023;43(2):163-175.
68. Smith DA, Sadler MC, Altman RB. Promises and challenges in pharmacoepigenetics. Camb Prism Precis Med. 2023;1:e18.
69. Ingelman-Sundberg M, Zhong XB, Hankinson O, et al. Potential role of epigenetic mechanisms in the regulation of drug metabolism and transport. Drug Metab Dispos. 2013;41(10):1725-1731.
70. Stefanska B, MacEwan DJ. Epigenetics and pharmacology. Br J Pharmacol. 2015;172(11):2701-2704.
71. Cascorbi I, Schwab M. Epigenetics in Drug Response. Clin Pharmacol Ther. 2016;99(5):468-470.

72. Ivanov M, Kacevska M, Ingelman-Sundberg M. Epigenomics and interindividual differences in drug response. Clin Pharmacol Ther. 2012;92(6):727-736.
73. Kelly TK, De Carvalho DD, Jones PA. Epigenetic modifications as therapeutic targets. Nat Biotechnol. 2010;28(10):1069-1078.
74. Ivanov M, Barragan I, Ingelman-Sundberg M. Epigenetic mechanisms of importance for drug treatment. Trends Pharmacol Sci. 2014;35(8):384-396.
75. Lötsch J, Schneider G, Reker D, et al. Common non-epigenetic drugs as epigenetic modulators. Trends Mol Med. 2013;19(12):742-753.
76. Wu Y, Shi X, Liu Y, et al. Histone deacetylase 1 is required for Carbamazepine-induced CYP3A4 expression. J Pharm Biomed Anal. 2012;58:78-82.
77. Zuk O, Schaffner SF, Samocha K, et al. Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A. 2014;111(4):E455-464.
78. Tasa T, Krebs K, Kals M, et al. Genetic variation in the Estonian population: pharmacogenomics study of adverse drug effects using electronic health records. Eur J Hum Genet. 2019;27(3):442-454.
79. Lin E, Lane HY. Machine learning and systems genomics approaches for multi-omics data. Biomark Res. 2017;5:2.
80. Athreya AP, Neavin D, Carrillo-Roa T, et al. Pharmacogenomics-Driven Prediction of Antidepressant Treatment Outcomes: A Machine-Learning Approach With Multi-trial Replication. Clin Pharmacol Ther. 2019;106(4):855-865.
81. van der Lee M, Swen JJ. Artificial intelligence in pharmacology research and practice. Clin Transl Sci. 2023;16(1):31-36.
82. Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. N Engl J Med. 2023;388(13):1201-1208.
83. Gottlieb A, Daneshjou R, DeGorter M, et al. Cohort-specific imputation of gene expression improves prediction of warfarin dose for African Americans. Genome Med. 2017;9(1):98.
84. Monaco A, Pantaleo E, Amoroso N, et al. A primer on machine learning techniques for genomic applications. Comput Struct Biotechnol J. 2021;19:4345-4359.
85. van der Lee M, Allard WG, Vossen R, et al. Toward predicting CYP2D6-mediated variable drug response from CYP2D6 gene sequencing data. Sci Transl Med. 2021;13(603).
86. Ortega VE, Meyers DA. Pharmacogenetics: implications of race and ethnicity on defining genetic profiles for personalized medicine. J Allergy Clin Immunol. 2014;133(1):16-26.
87. Zhou Y, Lauschke VM. The genetic landscape of major drug metabolizing cytochrome P450 genes-an updated analysis of population-scale sequencing data. Pharmacogenomics J. 2022; 22(5-6):284-293.
88. Patel JN. Cancer pharmacogenomics: implications on ethnic diversity and drug response. Pharmacogenet Genomics. 2015;25(5):223-230.
89. Zhang H, De T, Zhong Y, Perera MA. The Advantages and Challenges of Diversity in Pharmacogenomics: Can Minority Populations Bring Us Closer to Implementation? Clin Pharmacol Ther. 2019;106(2):338-349.
90. Perera MA, Cavallari LH, Johnson JA. Warfarin pharmacogenetics: an illustration of the importance of studies in minority populations. Clin Pharmacol Ther. 2014;95(3):242-244.

91. Nelson MR, Johnson T, Warren L, et al. The genetics of drug efficacy: opportunities and challenges. Nat Rev Genet. 2016;17(4):197-206.
92. Minderhoud C, Otten LS, Hilkens PHE, van den Broek MPH, Harmsze AM. Increased frequency of CYP2C19 loss-of-function alleles in clopidogrel-treated patients with recurrent cerebral ischemia. Br J Clin Pharmacol. 2022;88(7):3335-3340.
93. Lima JJ, Thomas CD, Barbarino J, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2C19 and Proton Pump Inhibitor Dosing. Clin Pharmacol Ther. 2021; 109(6):1417-1423.
94. Casolino R, Beer PA, Chakravarty D, et al. Interpreting and integrating genomic tests results in clinical cancer care: Overview and practical guidance. CA Cancer J Clin. 2024;74(3):264-285.
95. Klomp SD, Manson ML, Guchelaar HJ, Swen JJ. Phenoconversion of Cytochrome P450 Metabolism: A Systematic Review. J Clin Med. 2020;9(9).
96. Mostafa S, Polasek TM, Bousman CA, et al. Pharmacogenomics in psychiatry - the challenge of cytochrome P450 enzyme phenoconversion and solutions to assist precision dosing. Pharmacogenomics. 2022;23(15):857-867.
97. Klomp SD, Veringa A, Alffenaar JC, et al. Inflammation altered correlation between CYP2C19 genotype and CYP2C19 activity in patients receiving voriconazole. Clin Transl Sci. 2024; 17(7):e13887.
98. Kumar S, Sinha N, Gerth KA, Rahman MA, Yallapu MM, Midde NM. Specific packaging and circulation of cytochromes P450, especially 2E1 isozyme, in human plasma exosomes and their implications in cellular communications. Biochem Biophys Res Commun. 2017;491(3):675-680.
99. Rowland A, Ruanglertboon W, van Dyk M, et al. Plasma extracellular nanovesicle (exosome)-derived biomarkers for drug metabolism pathways: a novel approach to characterize variability in drug exposure. Br J Clin Pharmacol. 2019;85(1):216-226.
100. Koppel N, Maini Rekdal V, Balskus EP. Chemical transformation of xenobiotics by the human gut microbiota. Science. 2017;356(6344).
101. Bisanz JE, Spanogiannopoulos P, Pieper LM, Bustion AE, Turnbaugh PJ. How to Determine the Role of the Microbiome in Drug Disposition. Drug Metab Dispos. 2018;46(11):1588-1595.
102. Heirali A, Moossavi S, Arrieta MC, Coburn B. Principles and Terminology for Host-Microbiome-Drug Interactions. Open Forum Infect Dis. 2023;10(5):ofad195.
103. Alexander JL, Wilson ID, Teare J, Marchesi JR, Nicholson JK, Kinross JM. Gut microbiota modulation of chemotherapy efficacy and toxicity. Nat Rev Gastroenterol Hepatol. 2017; 14(6):356-365.
104. Xia L, Zhu X, Wang Y, Lu S. The gut microbiota improves the efficacy of immune-checkpoint inhibitor immunotherapy against tumors: From association to cause and effect. Cancer Lett. 2024;598:217123.
105. Zhang H, Hong Y, Wu T, et al. Role of gut microbiota in regulating immune checkpoint inhibitor therapy for glioblastoma. Front Immunol. 2024;15:1401967.
106. Trepka KR, Olson CA, Upadhyay V, Zhang C, Turnbaugh PJ. Pharma[e]cology: How the Gut Microbiome Contributes to Variations in Drug Response. Annu Rev Pharmacol Toxicol. 2025;65(1):355-373.