

From pharmacogenetics to pharmaco-omics: Milestones and future directions

Chiara Auwerx,^{1,2,3,4,5} Marie C. Sadler,^{2,3,4,5} Alexandre Reymond,^{1,*} and Zoltán Kutalik^{2,3,4,*}

Abstract

The origins of pharmacogenetics date back to the 1950s, when it was established that inter-individual differences in drug response are partially determined by genetic factors. Since then, pharmacogenetics has grown into its own field, motivated by the translation of identified gene-drug interactions into therapeutic applications. Despite numerous challenges ahead, our understanding of the human pharmacogenetic landscape has greatly improved thanks to the integration of tools originating from disciplines as diverse as biochemistry, molecular biology, statistics, and computer sciences. In this review, we discuss past, present, and future developments of pharmacogenetics methodology, focusing on three milestones: how early research established the genetic basis of drug responses, how technological progress made it possible to assess the full extent of pharmacological variants, and how multi-dimensional omics datasets can improve the identification, functional validation, and mechanistic understanding of the interplay between genes and drugs. We outline novel strategies to repurpose and integrate molecular and clinical data originating from biobanks to gain insights analogous to those obtained from randomized controlled trials. Emphasizing the importance of increased diversity, we envision future directions for the field that should pave the way to the clinical implementation of pharmacogenetics.

Introduction

Through the study of genetic determinants of drug responses, pharmacogenetics (PGx) plays a crucial role in achieving the promises of personalized medicine: providing a medical treatment tailored to one's genetic background. PGx variants modify a drug's pharmacokinetics by impacting its absorption, distribution, metabolism, and excretion (ADME), its pharmacodynamics by perturbing proteins involved in the drug's mechanism of action, or immune regulation.^{1,2} Understanding these interactions improves treatment outcome by optimizing drug dosage and efficacy while minimizing adverse drug reaction (ADR) risk.

To date, the Clinical Pharmacogenetics Implementation Consortium (CPIC)³ reports almost 450 gene-drug interactions. Eighty-three of these, involving 22 genes and 63 drugs, are annotated with the highest level of confidence and have prescription guidelines. These 63 drugs represent a small fraction of approved drugs but form a large portion of prescribed drugs,^{4,5} so that 35%–65% of the population has been exposed to at least one prescription drug with PGx indication.^{6–8} Notwithstanding, clinical implementation of PGx has been slow, with concerns regarding clinical validity and cost-effectiveness, infrastructure and data management, lack of awareness and education of health professionals, and ethical and regulatory issues being identified as the main barriers.⁹ To address these concerns, numerous PGx initiatives have been undertaken. So far, these revealed that pre-emptive genetic testing is cost effec-

tive in most situations, virtually benefits all given that >95% of individuals present at least one clinically actionable PGx variant, and could improve drug dosage while reducing ADRs.⁸

As encouraging results from these initiatives bring us a step closer to a widespread clinical implementation of PGx, this review takes a step back to discuss key milestones in the development of PGx methodology aiming at the identification, functional validation, and mechanistic understanding of clinically actionable germline variants (Figure 1). Purposefully omitting advances in oncology, where the consideration of somatic mutations adds an additional layer of complexity, we selected key studies that illustrate these milestones, highlighting how they have shaped PGx. The scientific community increasingly benefits from data streaming from high-throughput experiments, as well as biobanks coupling genetic information to intermediate molecular phenotypes and electronic health records. We emphasize remaining limitations and areas of active research, discussing strategies and methods that have not yet been fully developed and applied to PGx, envisioning the future of the field.

The origins of modern pharmacogenetics

The genetic basis of drug responses

The term “pharmacogenetics” was coined in 1959 by Friedrich Vogel¹⁰ to describe the emerging concept that drug

¹Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; ²Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; ³Swiss Institute of Bioinformatics, Lausanne, Switzerland; ⁴University Center for Primary Care and Public Health, Lausanne, Switzerland

⁵These authors contributed equally

*Correspondence: alexandre.reymond@unil.ch (A.R.), zoltan.kutalik@unil.ch (Z.K.)

<https://doi.org/10.1016/j.xhgg.2022.100100>.

© 2022 The Author(s). This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



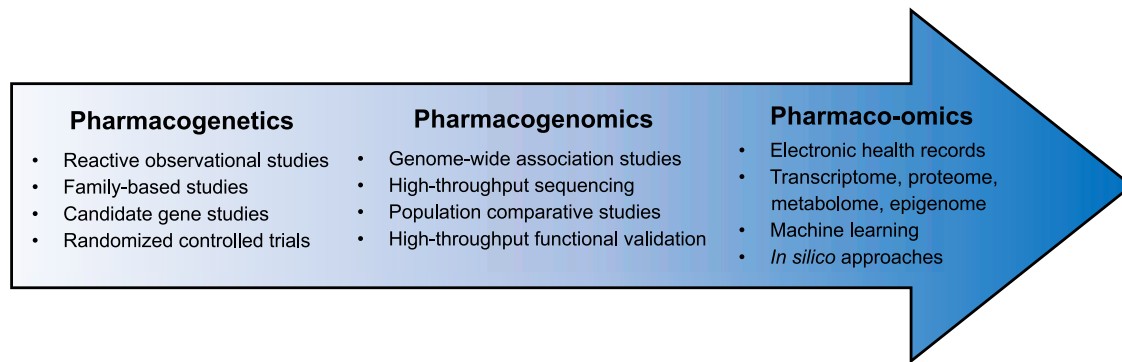


Figure 1. Overview of the chronological development of pharmacogenetics methodology

Each section of this review deals with a major PGx milestone: pharmacogenetics, pharmacogenomics, and pharmaco-omics. Listed are some of the main tools and approaches that were instrumental in the development of PGx at various stages. They will be discussed in their respective sections.

reactions are under genetic control.^{11,12} This idea originated from a series of observations indicating that individuals react differently when exposed to comparable amounts of a drug. For instance, ~10% of African Americans develop hemolytic anemia following treatment with the antimalarial drug primaquine, an ADR rarely observed in individuals of European ancestry (MIM: 300908, 305900).¹³ This observation was in line with a genetic origin of the ADR, as narrated by one of the founders of PGx, Arno Motulsky: “since a given gene may be more frequent in certain ethnic groups, any drug reaction that is more frequently observed in a given racial group, when other environmental variables are equal, will usually have a genetic basis.”¹² The intuition was confirmed decades later, as studies revealed a higher prevalence of causative glucose-6-phosphate dehydrogenase deficiency in Africans,¹⁴ tightly linking PGx with population genetics from the onset (Box 2).

Acknowledging the role of genetics in drug response promoted usage of classical human genetics tools to the study of PGx.²⁸ Twin studies could verify that a drug response was under genetic control by estimating its heritability, while family studies allowed inferring inheritance patterns of drug responses. For instance, inability to metabolize the antihypertensive drug debrisoquine²⁹ and the sodium channel blocker sparteine³⁰ follows an autosomal recessive pattern of inheritance (MIM: 608902). Inability to metabolize these compounds is strongly correlated and was believed to be controlled by the recently discovered cytochrome P450 (CYP) system.^{31,32} As oxidative metabolism of other drugs seemed unaffected,³¹ it suggested that only one enzyme of the CYP system was dysfunctional. Overall, these pioneering studies had the merit to confirm the underlying genetic basis of differential drug response, forming the foundations of PGx.

The candidate gene approach

A drawback of early PGx studies was their inability to map differential drug responses. Progress in molecular biology and improved understanding of cellular and physiological

processes set the stage for candidate gene studies (CGSs), which use prior knowledge to identify PGx interactions typically involving common or large-effect-size variants. After selecting candidate genes based on reported roles in drug ADME or relevant biological pathways, association between drug response and genetic variants within these genes is assessed.³³ For instance, the etiology of the debrisoquine and sparteine non-metabolizer phenotype was elucidated by purifying the relevant CYP³⁴ and cloning the relevant alleles.³⁵ This gene later came to be known as *CYP2D6* (MIM: 124030), one of the best studied PGx genes. Human *CYP2D6* is part of a family of 57 putatively functional genes, a dozen of which being responsible for the oxidative biotransformation of 70%–80% of all drugs,³⁶ arguably making them the most important PGx gene family. Other CGSs elucidated the genetic cause of abacavir hypersensitivity (MIM: 142830), a severe multisystemic response experienced by ~5% of human immunodeficiency virus (HIV)-infected patients treated with the antiviral drug.³⁷ Suspicion of a genetic origin for the ADR led two groups to assess variants in genes involved in abacavir metabolism and immune response, jointly identifying *HLA-B*5701* as the main genetic risk factor for abacavir hypersensitivity.^{38,39}

These success stories sparked the idea that new PGx knowledge could be brought into the clinic to optimize treatment. Given genotyping costs in the early 2000s,⁴⁰ it was crucial to validate interactions before adopting systematic screening. Most PGx interactions unraveled by CGS resulted from retrospective case-control studies comparing the frequency of suspected genetic variants between a group that experienced the drug response and a group of matched controls that did not.⁴¹ Despite being cost effective, their retrospective nature makes them prone to various types of biases. They also suffer from general challenges associated with epidemiological study design, such as improper definition of the outcome phenotype, lack of power due to inadequate sample size, or failure to control for stratification or pleiotropy.⁴¹ Hence, many associations failed to replicate and further proof was warranted prior to clinical implementation.

Box 1. Glossary

Allele: one of two or more versions of a genetic region.

Biomarker: measure of a biological state that is indicative of a physiological or pathological process.

Diplotype: combination of two haplotypes present in an individual.

Epistasis: phenomenon whereby the phenotypic consequence of a genetic variant is dependent on the presence of one or multiple other genetic variants (non-additive effects; gene-by-gene interactions).

Expression quantitative trait locus (eQTL): genetic locus harboring markers that explain a certain proportion of the variance in expression of a given gene.

Fine-mapping: statistical approach aiming at determining the most likely causal genetic variant(s) for an observed association signal.

Forward genetics: identification of the genetic basis of a phenotype.

Genetic architecture: underlying genetic basis of a phenotype (e.g., monogenic, polygenic, ...).

Genetic reference population: collection of strains from a model organism that were generated through specific breeding and for which broad genotype and phenotype data are available.

Genome-wide significance: threshold of significance that accounts for the fact that genome-wide association studies (GWASs) perform a large number of associations on not fully independent genetic variants. Typically, $p \leq 5 \times 10^{-8}$ is used.

Haplotype: group of alleles that are inherited together from a single parent.

Heritability: proportion of phenotypic variation in a given population that is due to genetic variation.

Imputation: inference of genetic variants that were not directly genotyped using haplotype reference panels. Imputation is possible as blocks of variants in linkage disequilibrium (LD) tend to be co-inherited, so that the presence of a given variant is predictive of the genotype at an adjacent location.

Linkage disequilibrium (LD): non-random association of alleles at two or more loci in a population.

Missense variant: SNV leading to the substitution of a different amino acid in the resulting protein.

Monogenic trait: trait influenced by genetic variants in or near a single gene.

Oligogenic trait: trait influenced by genetic variants in or near a few genes.

Pleiotropy: phenomenon whereby a single gene or variant affects multiple phenotypes.

Polygenic score (PGS): estimate of an individual's genetic liability to a trait calculated as the weighted sum of phenotype-associated variants, weights being the effect sizes from GWAS summary statistics.

Polygenic trait: trait influenced by genetic variants in or near a large number of genes.

Polypharmacy: co-prescription of several drugs.

Population differentiation: accumulation of differences in allele frequencies between populations with restricted gene flow due to natural selection or genetic drift.

Population stratification: presence of sub-populations with different allele frequencies within a seemingly homogeneous population. If not accounted for, population stratification can lead to spurious genotype-phenotype associations.

Reverse genetics: identification of the phenotypic consequence(s) of a particular genetic variant.

Sequencing coverage: number of unique reads that include a particular base pair.

Single nucleotide variant (SNV): substitution of a single nucleotide at a specific genomic location.

Star allele: allele following the standardized nomenclature for genetic polymorphisms in pharmacogenetics (PGx) genes. Star alleles are not restricted to single variants but can refer to functional haplotypes of PGx genes.

Summary statistics: aggregated GWAS results lacking individual genotype and phenotype data and typically including the variant's identifier, genomic location, reference and alternative allele, and their respective frequencies, a metrics for the strength of association with uncertainty and a p value.

Randomized controlled trials

Randomized controlled trials (RCTs) have become the gold standard of evidence-based medicine over the course of the 20th century.⁴² PGx RCTs allow measuring the effects of the genotype, treatment, and their interaction, providing causal evidence for the use of a genotype-informed treatment plan.⁴¹ Ideally, participants are randomized to receive either genotype-guided or conventional therapy (Figure 2A). Alternatively, when the genotype of interest is rare or when several treatments are to be compared, participants can be

stratified based on their genotype and randomized within each stratum (Figure 2B). Following two independent reports of *HLA-B*5701* as the main genetic risk factor for abacavir hypersensitivity,^{38,39} a prospective RCT set out to investigate the clinical benefits of *HLA-B*5701* screening by randomizing ~2,000 HIV-positive patients according to the design of Figure 2A.⁴³ While all participants in the control group received abacavir, individuals in the intervention group were genotyped so that only *HLA-B*5701*-negative patients received the drug. The study demonstrated that

Box 2. Pharmacogenetics for all

Why is diversity relevant?

Ability to assess the phenotypic impact of a variant depends on its frequency in the studied population. Early on, it was observed that different populations do not respond equally to certain drugs.¹³ Later, microarray studies revealed population differentiation across ADME genes^{15,16} and sequencing studies generated global maps of PGx allele frequencies.^{17,18} Following, a gene-based population differentiation method that optimally accounts for rare variants showed that 10% of investigated PGx genes exhibit high population differentiation.¹⁹

What are the consequences of lack of diversity?

Despite growing awareness around the importance of inclusiveness, the majority of genetic studies remain Eurocentric, undermining the ability to discover new PGx variants and implement equitable clinical guidelines. A 2021 systematic review showed that 88% of discovery PGx genome-wide association studies (GWASs) participants were of European ancestry.²⁰ This bias is apparent in the largest PGx study to date: 17% of the UK Biobank participants carry at least one undocumented deleterious variant in the 14 analyzed PGx genes, the majority of which being of non-European ancestry.²¹ Similarly, 72% of the participants enrolled in clinical trials leading to US Food and Drug Administration (FDA) drug approval in 2019 were of white ancestry.²² Biased and underpowered, such studies miss or underestimate the strength of robust PGx interactions involving variants predominantly observed in minority populations. This results in poor generalizability of findings, which translate into healthcare disparities.²³ For example, polygenic scores (PGSs) ([Polygenic pharmacogenomics](#)) ought to play a major role in personalized medicine but raise ethical concerns due to poor accuracy in individuals of non-European ancestry.²⁴

How can diversity be leveraged to enable new PGx discoveries?

Studies discussed in this review illustrate how increased diversity can benefit PGx. Including participants of both European and African ancestry, it was shown that the strength of association between hepatitis C treatment efficacy and *IFNL3* genotype was comparable in both populations and that the increased prevalence of the unfavorable *IFNL3* genotype in Africans explained a large fraction of the differences in treatment responsiveness across groups ([The first pharmacogenomics GWAS](#)).²⁵ In another example, three out of five regions associated with bronchodilators response in asthmatic children were discovered by analyzing minority populations. However, these proved difficult to validate due to the absence of replication cohorts ([The human pharmacogenomic landscape](#)).²⁶ A third study proposes a mechanistic model to improve warfarin dose prediction by analyzing both European and African American cohorts ([Molecular modelling of pharmaco-omics interactions](#)).²⁷ Trained on each ancestry group, the model was used to predict maintenance dose in an ancestry-matched and cross-ancestry manner, yielding improved results compared with existing algorithms. This study highlights how a mechanistic approach to modeling combined with a limited number of samples of different ancestries allows to broaden model applicability to other ethnic groups. Overall, inclusiveness can be applied to a wide range of methodologies, leading to knowledge that benefits all.

pre-emptive screening reduced the incidence of clinically diagnosed abacavir hypersensitivity from 7.8% to 3.4% ($p < 0.001$) and eliminated immunologically confirmed abacavir hypersensitivities (2.7% in controls; $p < 0.001$),⁴³ prompting the establishment of clinical screening guidelines.⁴⁴ Recently, the impact of these guidelines was evaluated, showing that, since their introduction, the proportion of patients undergoing pre-emptive screening steadily increased, while ADR incidence decreased 6-fold.⁴⁵

RCTs provide strong support of causality between an exposure and an outcome, but their role in PGx has been called into question.^{23,46} RCTs present multiple limitations, often related to their high costs. Systematic investigation of 138 clinical trials estimated the median cost of a trial to \$19 million,⁴⁷ with costs increasing with participant number and the inclusion of a randomized control group. Studying rare PGx variants or drug response requires large sample sizes and quickly becomes financially unsustainable. This problem is illustrated with the TAILOR-PCI RCT that found no statistically significant reduction in major adverse cardiovascular events (MACEs)

(12 months follow-up) upon *CYP2C19*-guided prescription of anti-platelet medication (hazard ratio = 0.66; 95% confidence interval [CI]: 0.43–1.02; $p = 0.06$).⁴⁸ Clopidogrel is the most widely prescribed anti-platelet medication,⁴⁹ but due to genetically driven variability in drug response, it is suspected to increase the risk for MACEs.⁵⁰ As *CYP2C19* (MIM: 124020, 609535) is required to activate clopidogrel,⁵¹ *CYP2C19* loss-of-function (LOF) carriers in the genotype-guided group were prescribed ticagrelor instead of clopidogrel.⁴⁸ Despite its large size ($n = 5,302$), the trial was only powered (85%) to detect a minimum hazard ratio of 0.50, so that the detected trend did not reach significance.⁴⁸ Another limitation is that only two LOF alleles were considered, although recent evidence demonstrated a polygenic architecture of clopidogrel response ([Polygenic pharmacogenomics](#)).⁵² Considering these results, the US Food and Drug Administration (FDA) does not require genetic testing prior to clopidogrel treatment initiation, nor do the American College of Cardiology and European Society of Cardiology recommend routine *CYP2C19* testing. Notwithstanding, recent meta-analyses suggest

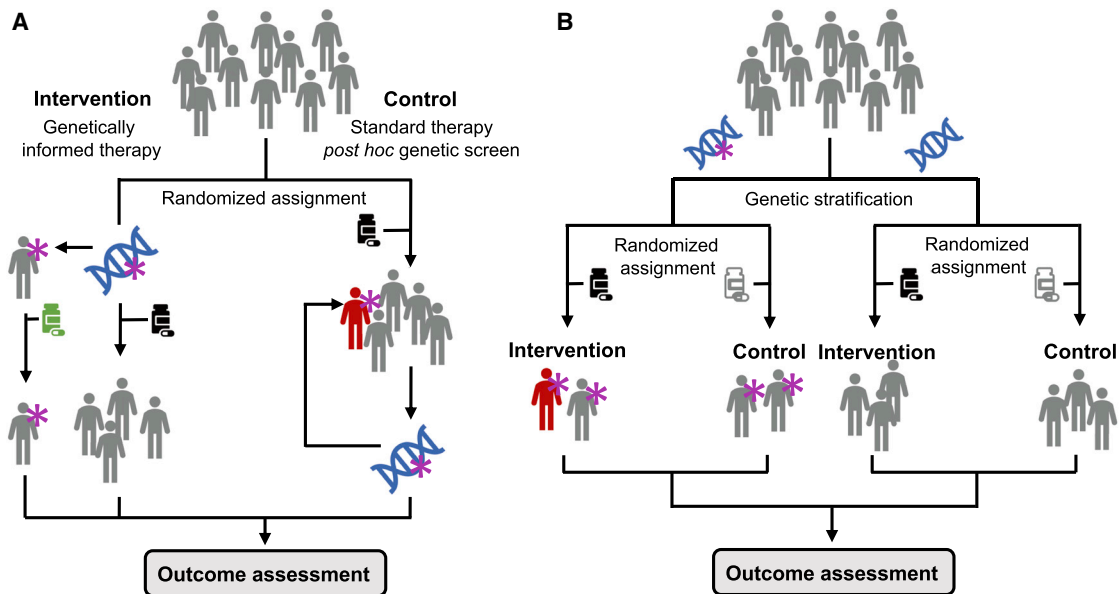


Figure 2. Possible study designs for pharmacogenetic randomized controlled trials

(A) Participants are randomly assigned to an intervention or control group. In the former, participants are screened for the presence of the PGx variant of interest (star). Negative individuals receive standard treatment (black), while positive ones receive an adapted alternative treatment (green). In the control group, all individuals receive conventional treatment and genetic screening is performed *post hoc*. The number of participants exhibiting the response of interest (red) is assessed and compared among groups.

(B) Participants are stratified based on the presence or absence of the PGx variant of interest. In each strata, participants are randomly assigned to an intervention group, which receives standard treatment (black) and a control group, which receives a placebo or alternative treatment (white). The number of participants exhibiting the response of interest is assessed and compared among groups and strata.

that *CYP2C19*-guided prescription could identify patients benefiting from alternative anti-platelet medication,^{53,54} prompting the CPIC to update its recommendations in 2022, thereby providing guidance to clinicians on how to interpret *CYP2C19* genetic test results.⁵⁵ Concurrently, a subset of the TAILOR-PCI RCT has increased the follow-up period to 24 months, additionally performing a feasibility study to assess digital solutions allowing more efficient and cost-effective patient follow-up and study design.⁵⁶ Besides the costs and hurdles associated with generating prospective evidence of clinical utility, there is a lack of incentive for the pharmaceutical industry to conduct PGx RCTs, as patents for most drugs with PGx guidelines—including clopidogrel—have expired, making them unprofitable.^{2,23} Another recurrent issue is the lack of diversity (Box 2). Finally, PGx RCTs are associated with serious ethical concerns: Rarely performed without prior knowledge, some study designs (e.g., Figure 2B) expose patients with a likely actionable variant to a treatment that puts them at risk of ADRs.^{23,41}

Slow clinical implementation of PGx is often credited to lack of supporting evidence by RCTs, which can partially be attributed to the limitations described above. RCTs remain an important epidemiological tool, as demonstrated by the more than 800 PGx trials reported over the last 20 years, of which 155 are still recruiting or ongoing. However, RCTs should not be the sole approach to validate PGx interactions. Proof of efficacy of genetic testing often needs to meet higher standards compared with non-ge-

netic tests, and most physicians unconsciously practice PGx by adjusting patient's treatments.⁴⁶ To ensure transfer of PGx knowledge from bench to bedside while maintaining rigorous standards of clinical efficacy, it will be important to find more efficient ways to conduct RCTs and broaden the spectrum of accepted proof of evidence. In [Leveraging observational data to gain causal insights](#), we introduce statistical tools leveraging large-scale observational data to approximate the effects of RCT studies.

The genomic revolution

The pharmacogenome

Sequencing of the human genome in the early 2000s^{57,58} brought forward the concept of the druggable genome, originally defined as genes whose product can or are predicted to be bound by small molecules.⁵⁹ Besides providing a blueprint to map genes and variants associated with drug responses, access to the reference genome catalyzed waves of technological progress in genomics (Box 3). A common theme is the possibility to assess more—and potentially novel—variants with higher accuracy and at ever lower prices, leading to the metamorphosis of *pharmacogenetics* into *pharmacogenomics*. Conceptually similar and therefore both referred to as PGx hereafter, pharmacogenomics aims at discovering genetic variants affecting drug response by probing the entire genome, as opposed to a few candidate loci. This progress enabled the exploration

Box 3. Overview of genomic technologies

Microarray genotyping: Genomic variability can be assessed with genotyping microarrays that probe the identity of 250,000–4,000,000 predefined single nucleotide variants (SNVs).⁶⁰ Diverse, high-coverage, haplotype reference panels are used to impute variants at positions that were not genotyped. In spite of increasing the multiple-testing burden, boosting the number of assessed SNVs increases power, improves fine-mapping, and facilitates meta-analyses.⁶¹ Imputed microarray data allow the assessment of millions of SNVs for a fraction of sequencing costs.

Short-read sequencing: A major breakthrough came from the development of high-throughput next-generation sequencing (NGS) methods that implement massively parallel sequencing of short DNA fragments.⁶² Applications include whole-genome sequencing (WGS), which assesses the entire genome, and whole-exome sequencing (WES), which uses capture and enrichment methods to selectively sequence exons, the protein coding regions of the genome. Because exons represent ~ 1% to 2% of the human genome,⁵⁸ for the same amount of sequencing, WES can increase coverage or sample number. Cost effective and better suited to assess rare variants and diverse populations, low coverage sequencing (~ 1×) combined to imputation represents a promising alternative to microarrays.^{63,64}

Long-read sequencing: Bypassing the need for polymerase chain reaction, third-generation sequencing based on long reads reduces the risk of amplification-related errors and biases. By sequencing DNA fragments over 1 Mb long,^{65,66} this technology can accurately resolve complex regions, call haplotypes, and map structural variants (e.g., copy-number variants).⁶⁷ Long-read sequencing can also detect base modifications,⁶⁸ providing further insights into the analyzed sequence.

CRISPR-based genome editing: CRISPR technologies allow to assess the impact of precise genetic modifications, ranging from point mutations^{69,70} to whole-chromosome deletions,^{71,72} in physiological settings. CRISPR tools are available for most model organisms, cell lines, and patient-derived cells, with a growing number of functionalities extending beyond simple gene editing.^{73–75} It is possible to generate whole-body or tissue-specific gene knockins and knockouts, combine both approaches to create humanized models by excising the endogenous gene and replacing it by its human version, or perform epigenetic editing. Several alterations can be induced simultaneously, opening the door to the study of epistatic interactions.

of the pharmacogenome, which we define as regions of the human genome that influence response to medication.

For PGx to embrace the genomic revolution, ways to effectively organize, store, and share newly generated knowledge were required, leading to the creation of PGx databases (Table 1). The further the field advances, the more it will rely on the presence of centralized, well-structured, and inter-connected resources. Search terms *pharmacogenetics/pharmacogenomics* returned over 30,000 PubMed entries published between 2000 and 2020, testifying how the genomic revolution allowed the collection of data and knowledge at an unprecedented scale.

Genome-wide association studies

Genome-wide association studies (GWASs) are observational studies wherein a large set of genetic variants spread across the genome, typically single nucleotide variants (SNVs) assessed with microarrays (Box 3), is scanned to identify associations with a phenotype. Hence, GWASs mitigate biases inherent to CGSs, enabling data-driven discovery of unsuspected PGx variants.

The first pharmacogenomics GWASs

One of the first PGx GWASs found that each additional copy of the *SLCO1B1* rs4149056 T>C allele increased the odds ratio for statin-induced myopathy by 4.5 (95% CI: 2.6–7.7; $p = 2 \times 10^{-9}$; MIM: 604843).⁸⁶ If the gene was known to encode a hepatic statin transporter,⁸⁷ it had

not been linked to the ADR. Due to the rarity of the ADR, the GWAS was conducted within the larger SEARCH RCT, which followed ~12,000 participants exposed to different doses of statins over the course of 6 years, providing access to matched controls. Case number was doubled by considering both individuals with definite and milder myopathy, highlighting the importance of adequate phenotype definition. These steps were crucial in gathering enough participants to have the statistical power to detect associations. The challenge is exacerbated in PGx studies, as only a fraction of individuals take the drug of interest and even a smaller fraction might develop a given ADR. Review of 23 PGx GWASs conducted between 2007 and 2010 showed that, on average, drug-response GWASs had 570 participants and ADR GWASs a case/control ratio of 70/120.⁸⁸ Estimates from 2015 to 2020 show that these numbers are increasing (median: 1,220 participants) but remain ~10 times lower than for non-PGx GWASs.²⁰ An indirect consequence is that fewer GWASs have investigated drug responses.^{20,89} To overcome this hurdle, PGx GWASs can be performed within the context of RCTs or conducted by international consortia (e.g., PGRN-RIKEN),⁹⁰ facilitating the recruitment of large cohorts.

So far, most examples focused on ADRs, which are easy to identify but rare. Conversely, providing sufficient technical and financial means, drug response can be assessed in all treated subjects, offering greater statistical power to

Table 1. List of five major pharmacogenetics knowledge databases and resources

Name	Aim	Reference and website
CPIC: The Clinical Pharmacogenetics Implementation Consortium	facilitate the clinical implementation of PGx by generating curated, evidence-based, and updated guidelines that provide prescription recommendations for gene-drug pairs	Relling and Klein ⁷⁶ , Relling et al. ³ , https://cpicpgx.org/
DGIdb: The Drug Gene Interaction database	bridge the gap between drug discovery and PGx by integrating data from other databases, such as PharmGKB, Therapeutic Target Database, ⁷⁷ DrugBank, ⁷⁸ etc.	Griffith et al. ⁷⁹ , Freshour et al. ⁸⁰ , https://www.dgidb.org/
PGRN: The Pharmacogenomics Global Research Network	PGx hub with the aims of <ul style="list-style-type: none"> ● providing a platform for the PGx community ● promoting and advancing research in PGx ● bringing awareness to the importance of PGx Coordinates PharmGKB, CPIC, and PharmVar	Relling et al. ⁸¹ , https://www.pgrn.org/
PharmGKB: The Pharmacogenetics Knowledge Base	first centralized gene-drug interaction database aiming at linking genomic data to molecular and cellular phenotypes, as well as to clinical information	Whirl-Carrillo et al. ⁸² , Whirl-Carrillo et al. ⁸³ https://www.pharmgkb.org/
PharmVar: The Pharmacogenetic Variation Consortium	provide a centralized repository for all PGx variants and a standardized nomenclature for PGx alleles	Gaedigk et al. ⁸⁴ , Gaedigk et al. ⁸⁵ , https://www.pharmvar.org/

detect PGx interactions. One of the first drug responses investigated by GWAS was warfarin maintenance dose.^{91,92} At the time, warfarin was the most widely used oral anticoagulant, with prescription rates totaling 0.5%–1.5% of the population.⁹³ The therapeutic window of warfarin is narrow and exhibits inter-individual variability, about 50% of which being explained by patient-specific demographics, clinical factors, and genetic background.⁸⁸ Based on prior knowledge of warfarin metabolism (Figure 3), CGSs demonstrated the independent role of polymorphisms in *VKORC1* (MIM: 608547) and *CYP2C9* (MIM: 601130) in determining an individual's drug dosage and ADR risk due to drug concentrations outside the therapeutic index.⁹⁴ To determine whether other genes were implicated, a GWAS on warfarin dosage was conducted (n = 181), identifying a single genome-wide significant signal near *VKORC1*.⁹¹ A second GWAS with improved power (n = 1,053) identified significant signals for *VKORC1*, *CYP2C9*, and *CYP4F2* (MIM: 604426).⁹² The fact that these genome-wide studies did not reveal new genetic risk factors reflects that candidate genes with mechanistic support are bound to have strong PGx effect and will have the highest power to be detected by GWASs. Nevertheless, they corroborated earlier findings through an orthogonal method, excluding the presence of additional large-effect variants.

GWASs also discovered new PGx interactions. Before 2011, hepatitis C was treated with a poorly tolerated combination of pegylated interferons and ribavirin that failed to clear the infection in ~50% of patients.⁹⁵ Differential success rate across populations suggested a genetic origin for treatment efficacy (Box 2), prompting analysis through GWASs and revealing an association signal near *IFNL3* (previously *IL28B*; MIM: 607402, 609532).^{25,96–98} Arrival of direct-acting antivirals improved treatment success rate, especially in patients with unfavorable *IFNL3* geno-

type.⁹⁹ Nowadays, new generations of well-tolerated and highly effective direct-acting antiviral regimens bypass the need for pegylated interferons altogether.¹⁰⁰ This highlights an alternative to PGx screening: identifying drugs that elicit a more homogeneous response. Not always available, this alternative can lead to other issues, as new antivirals are associated with drug-drug interactions,¹⁰¹ themselves influenced by PGx. Together, GWASs are an important complement to hypothesis-driven CGSs and have discovered or replicated 586 unique drug-variant associations in the 14 years since the first published PGx GWASs.²⁰

Polygenic pharmacogenomics

A long unanswered question relates to the genetic architecture of PGx traits. In 1918, R.A. Fisher showed that continuous traits can be explained through Mendelian inheritance of many genetic variants with a genuine, albeit small, effect.¹⁰² Since, the polygenic nature of complex traits has become widely accepted and pushed to the extreme with the omnigenic model, which hypothesizes that all genes expressed within a phenotype-relevant cell type contribute to the phenotype in question.^{103,104} Despite the quantitative nature of drug responses, early GWASs pointed to a mono- or oligogeneic architecture governed by few large-effect-size variants detectable by CGSs or small GWASs. Demonstrating polygenicity in PGx traits has not been straightforward. In the 2000s, CGSs implicated several genes in the response to the anti-diabetic drug metformin,^{105–108} but these loci were not sufficient to explain the estimated 34% heritability of the drug response.¹⁰⁹ Instead, chromosome-wise heritability estimates pointed at the presence of multiple small-effect variants spread across the genome.¹⁰⁹ In the following years, independent, better-powered GWASs implicated an increasing number of genes in metformin response,^{110–113}

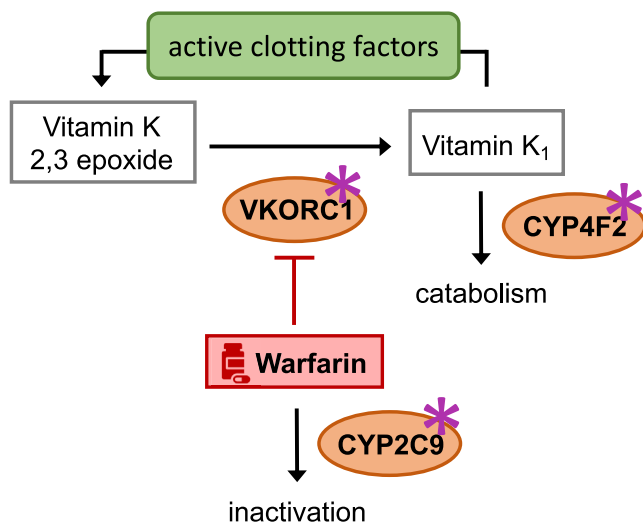


Figure 3. Warfarin pharmacogenetics

Main enzymes (orange) encoded by genes whose polymorphisms (star) affect warfarin (red) dosage. Warfarin inhibits VKORC1, an enzyme that reduces vitamin K₁ 2,3 epoxide to vitamin K, so that the latter can act as a cofactor for clotting factor activation. CYP2C9 is the main enzyme involved in the inactivation of warfarin. CYP4F2 is involved in vitamin K₁ catabolism. Metabolites are in square boxes.

supporting the predicted polygenic architecture. More recently, heritability of 12 PGx traits was estimated between 5% and 59%, with 62%–95% of the heritability being attributed to moderate- and small-effect variants.¹¹⁴ These results suggest a highly polygenic architecture for these traits, indicating that larger GWASs are likely to reveal further associations.

Improving genotype-based patient classification, polygenic scores (PGSs) have gained popularity to estimate the cumulative effect of genome-wide genetic variation on a phenotype. Given their success in identifying individuals at risk for common diseases,¹¹⁵ it is of interest to determine whether PGSs can help predict an individual's drug response. So far, applications in PGx are sparse and have yielded mixed results, as highlighted in a 2021 comprehensive review of 51 PGx PGS studies.¹¹⁶ One of the most widely studied trait is response to antidepressants. Despite being an excellent candidate trait given the large number of treatment-refractory patients¹¹⁷ and a high heritability,¹¹⁸ only recently was its polygenic basis confirmed.^{118,119} Other studies started demonstrating the usefulness of PGx PGSs. Prior knowledge was used to design a PGS for ADR risk to the previously discussed anti-platelet drug clopidogrel.⁵² Six SNVs associating with on-treatment platelet reactivity were used to build a predictive PGS for MACEs, despite none of the variants associating with the trait, thereby demonstrating the predictive value of small cumulative effects. Another study developed a series of PGSs for drug-induced liver injury (DILI),¹²⁰ a rare and genetically determined ADR induced by a wide range of drugs.^{121,122} In addition to validation in an independent dataset, PGSs were tested in liver

organoids and primary hepatocytes, revealing shared predictivity across different drugs. Cell-based and transcriptomics assays further provided mechanistic insights. Free from the ethical and technical considerations associated with human studies, this “polygenicity-in-a-dish” approach allows to investigate the downstream impact of diverse genetic profiles in a tissue-specific and environmentally controlled way. Similar approaches with tailored follow-up assays might be applied to study other PGx traits.

Beyond classical challenges (sample sizes, phenotype definition, integration of rare variants, etc.), limitations intrinsic to microarray data will have to be overcome to fully understand the genetic architecture of PGx traits. It is difficult to design probes for the highly polymorphic, repetitive, or complex regions many PGx-relevant loci map to (e.g., *CYPs* and *HLA*). Though newer arrays include more PGx content, they do not yet provide perfect coverage of known PGx alleles.⁶⁰ Furthermore, microarrays designed to assess common variants in Europeans provide suboptimal coverage in other populations (Box 2). This is mitigated by new arrays optimized for population-specific or multiethnic studies and ever larger and more diverse reference panels (e.g., TOPMed Imputation Reference panel)¹²³ that improve imputation quality across diverse ancestries. Lastly, microarrays are not suited to accurately assess rare variants,^{124,125} despite evidence suggesting that these might account for some of the missing heritability in PGx traits.¹²⁶

Next-generation sequencing in pharmacogenomics

Unlike genotyping, sequencing assesses the exact order and identity of all nucleotides within a DNA segment (Box 3), allowing the discovery of unreported genetic variants. With decreasing sequencing price,⁴⁰ it becomes possible to (1) increase sequencing coverage, leading to improved accuracy, and (2) sequence individuals from diverse populations, leading to the identification of new genetic variants.

The human pharmacogenomic landscape

Next-generation sequencing (NGS) has widely been used to gauge the mutational landscape of known PGx genes. If non-targeted approaches, such as whole-exome sequencing (WES) or whole-genome sequencing (WGS) (Box 3) allow the unbiased assessment of the mutational spectrum among all human genes, resources can be optimized by evaluating with increased coverage a set of genes likely to yield new insights. For instance, PGRNseq, a capture panel that sequences 84 PGx genes with ultra-deep 500× coverage,¹²⁷ will assess PGx variation in the electronic Medical Records and Genomics (eMERGE)-PGx project (n ≈ 9,000).^{128,129} Another study (n = 14,002) analyzed rare variants in 202 drug target genes, suggesting that rare variants are both numerous and population specific.¹³⁰ Genetic characterization of the 57 human *CYP* genes revealed over 6,000 new SNVs.¹³¹ Similar studies sampling individuals from various ethnicities were later conducted for a broader spectrum of PGx genes, coming

Box 4. Rare variant association testing in pharmacogenomics

Burden tests collapse rare variants within a predefined region or gene into a genetic score. As they falsely assume the same direction and magnitude of effect for all grouped variants, they can be inaccurate. For instance, variants in *CYP2C9* can increase, decrease, or abolish *CYP2C9*'s ability to metabolize warfarin.¹³⁵

Variance-component tests investigate rare and common variants within a region simultaneously and have increased power in the presence of opposite effect variants or when only a small fraction of the variants is causal. This comes at the cost of requiring larger sample sizes, which can be challenging in PGx. The best known variance-component test is the sequence kernel association test (SKAT).¹³⁶

Combined tests combine features from both above-mentioned tests, increasing robustness and performance in small sample sizes. The most widely used is the optimal unified SKAT test (SKAT-O).¹³⁷

to the conclusion that >90% of SNVs in these genes are rare.^{132,133} Despite their low frequency, rare variants strongly contribute to functional variability,^{126,132} many being predicted to be deleterious.¹⁷

At fixed effect size, sample size required to detect an association increases with decreasing allele frequency, so that GWASs lack the power to detect rare variant associations. This led to the development of statistical methods for cumulative rare variant association testing¹³⁴ (Box 4), but few PGx studies have made use of them so far. Recently, optimal unified sequence kernel association test (SKAT-O) was used on WGS data to identify genetic determinants for bronchodilators response in 1,441 asthmatic children of diverse ethnicities (Box 2).²⁶ Five significant SKAT-O regions were identified at non-ADME loci, explaining 4%–8% of the phenotypic variation. This exemplifies how new PGx discoveries can be made by promoting diversity and investigating rare variants in genes lacking prior known PGx interactions.

The non-coding pharmacogenome

Despite most GWAS signals mapping to non-coding regions, fine-mapping and mechanistic interpretation of these signals remains difficult.¹³⁸ In the previously discussed association between *SLCO1B1* and statin-induced myopathy, the variant originally identified by the GWAS was intronic. Only after linkage disequilibrium (LD) analysis was the likely causal missense variant, rs4149056, identified.⁸⁶ PGx associations mapping to non-coding regions could be explained by synthetic associations, a concept referring to the apparent association of a common marker with a trait resulting from that marker being in LD with one or multiple unobserved, rare, causal variants.¹³⁹ Drug-response-associated non-coding variants could simply tag the deleterious mutational burden in corresponding PGx genes,¹⁴⁰ even if it has been argued that synthetic associations created by rare variants are unlikely to explain most GWAS results.¹⁴¹

Non-coding variants can mechanistically influence drug response, e.g., by influencing the binding of drug-regulated transcription factors. This concept is demonstrated by a variant (rs4743771 C>A) affecting the response to rosiglitazone,¹⁴² an antidiabetic drug associated with adverse effects, such as increased cholesterol levels.¹⁴³ Rosiglitazone medi-

ates its therapeutic effect by binding and activating the nuclear receptor PPAR γ (MIM: 601487),¹⁴⁴ which induces the expression of multiple genes, including the cholesterol efflux transporter *ABCA1* (MIM: 600046).¹⁴⁵ Chromatin immunoprecipitation sequencing showed that A/A carriers lack a rosiglitazone-induced PPAR γ -binding site and fail to induce expression of nearby *ABCA1* upon administration of the drug.¹⁴² Confirming the clinical relevance of the variant, diabetic patients with the corresponding genotype experienced a reduced rosiglitazone-induced increase in cholesterol levels.¹⁴² Overall, combining NGS to biochemical and molecular assays can shed light on PGx variants in regulatory regions. Another emerging mechanism of action is via long non-coding RNAs (lncRNAs) modulation.¹⁴⁶ While this area of research remains in its infancy, it was shown that the transcription factor HNF1 α (MIM: 142410) and the lncRNA HNF1 α -AS1 form a regulatory network that controls the expression of several *CYP* genes.^{147,148}

Elucidating complex pharmacogenomic regions

Obtaining fine-scaled sequencing of PGx genes residing in complex genomic regions remains challenging. Previously mentioned *CYP2D6* is notoriously difficult to analyze with NGS due to the presence of homologous pseudogenes and a propensity for structural rearrangements.¹⁴⁹ The gene is also highly polymorphic, with 149 alleles listed on PharmVar. Long-read sequencing workflows (Box 3) have been implemented to resolve *CYP2D6* haplotypes and diplotypes.^{150–154} Besides recovering known polymorphisms, these studies report a large fraction of new variants. As *CYP2D6* metabolizes ~25% of commonly prescribed drugs and that *CYP2D6* genotype directly influences drug metabolism efficiency,¹⁴⁹ obtaining accurate diplotypes could improve *CYP2D6* metabolizer phenotype prediction. While long-read sequencing technologies offer exciting perspectives and first population-scale applications are emerging,¹⁵⁵ they remain expensive and limited by low throughput and high error rates.

Experimental pharmacogenomics

Experimental settings provide precise control over genetic and environmental factors (e.g., diet, lifestyle, medication, and combinations thereof) and reduce the influence of confounders,¹⁵⁶ making it possible to establish causal

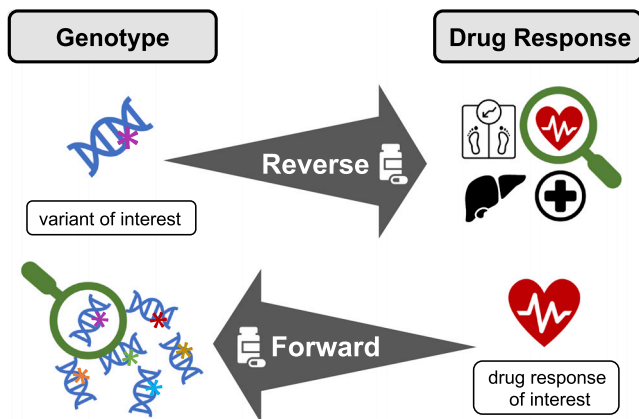


Figure 4. Schematic representation of two experimental approaches to pharmacogenetics

The reverse genetics approach starts with a genetic variant of interest and aims at characterizing its phenotypic consequences upon drug exposure. Conversely, the forward genetics approach starts with a phenotype of interest and screens a genetically diverse population to identify genes or variants eliciting the studied response upon drug exposure.

genotype-phenotype relationships and study complex phenomena, such as gene-by-gene (i.e., epistasis), gene-by-environment, and gene-by-sex interactions. The importance of understanding these effects is highlighted by an epidemiological study that found female sex and polypharmacy as the strongest predisposing factors for ADRs.¹⁵⁷ Pharmacological differences between sexes have long been acknowledged, and there is emerging evidence that these might influence PGx.^{158,159} Moreover, polypharmacy might impact ADR risk through drug-drug interactions¹⁶⁰ and drug-drug-gene interactions.¹⁶¹ Model organisms offer a convenient setting to disentangle these complex interactions through two experimental strategies (Figure 4).

A reverse genetics approach to pharmacogenetics

Reverse genetics investigates the phenotypic impact of a specific polymorphism by itself or in different sexes, genetic backgrounds, tissues, and/or environments. Before 2012, reverse genetics relied on laborious cloning techniques. Since, clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein 9 (Cas9)¹⁶² revolutionized genetic engineering through its simplicity, versatility, and precision (Box 3) and over 50 publications already made use of CRISPR to induce full PGx gene knockouts or characterize specific variants.¹⁶³ For instance, CRISPR-induced deletion of two linked SNVs in an upstream enhancer of *CYP2D6* revealed which one was responsible for increasing the expression of the gene.^{164,165} CRISPR technologies have been linked to off-target mutagenesis at sites with high sequence similarity¹⁶⁶—which are particularly common among ADME genes¹⁶⁷—even if careful usage of the tool should mitigate this issue.¹⁶⁸ More importantly, the large number of PGx variants and persistent increase in drug prescription and

polypharmacy¹⁶⁹ render it impossible to assess the full combinatorial space of experimental settings at current experimental throughput, limiting usage of CRISPR to strong candidate loci.

A forward genetics approach to pharmacogenetics

Forward genetics exploits genetically diverse population maintained under controlled environmental conditions to facilitate the identification of genetic modifiers of drug response. Contrasting with targeted approach, CRISPR technologies can also be used to perform high-throughput gene-level genome-wide knockout, knockdown, or knockin screens to assess the low-dimension combinatorial effect of genetic perturbations. Screens can be coupled to high-throughput molecular techniques, such as bulk or single-cell RNA sequencing, allowing measurement of downstream consequences of induced mutations.⁷⁴ Since drug resistance or susceptibility measured as cell proliferation or death can easily be assessed and exhibits limited tissue specificity, CRISPR screens have primarily been applied to cancer PGx research.¹⁷⁰ In the future, screens could characterize a wider array of PGx-relevant readouts and move toward the investigation of specific SNVs. Deep mutational scanning takes a step in this direction by probing the function of thousands of variants in a single experiment¹⁷¹ and was recently applied to generate abundance and activity scores for 87% of the possible missense variants in *CYP2C9*.¹⁷² Providing unprecedented functional insights into the mutational landscape of the gene, these data can be used to train functional prediction algorithms (Functional prediction of rare variants). Another source of genetic diversity are genetic reference populations, which are available for most model organisms (e.g., mouse).¹⁷³ Producing genetically identical individuals, they represent powerful mapping tools allowing the *in vivo* multi-tissue characterization of drug responses and can be coupled to extensive phenotyping and follow-up studies.¹⁷⁴ An *in vitro* alternative that directly mimics human genetic diversity is to harvest patient cells, which can be studied as primary cells or induced into pluripotent stem cells to generate organoids, as in the previously described drug-induced liver injury PGS study.¹²⁰ Patient-derived organoids can mimic an increasing number of organs and tissues, rendering the technology apt for personalized medicine strategies aiming at selecting the most efficient and safe treatment.¹⁷⁵ While widespread clinical implementation awaits, it was shown that organoids from cystic fibrosis patients can predict drug efficacy.^{176,177}

By studying appropriate and genetically diverse model systems, questions that were unanswerable due to ethical, logistical, or technical concerns become solvable. Experimental studies have their limitations, such as high costs, limited throughput, and/or poor reproducibility and translatability,^{178,179} but the virtuous cycle between human observational studies and *in vitro* and *in vivo* experimentation in model organisms¹⁵⁶ should incite PGx to exploit the respective strengths of these approaches.

Pharmacogenomics in the era of big data

Electronic health records and biobanks

Electronic health records (EHRs) have become widespread over the past decade as, besides supporting and improving diagnosis, clinical decisions, and treatment coordination, they provide new data analytics opportunities.¹⁸⁰ EHRs commonly encompass patient demographics, medical history, drug prescriptions, and in some cases laboratory results, radiological images, and wearable device data.¹⁸¹ Integration of EHRs is key to the realization of precision medicine, and current research already benefits from it by conducting nonrandomized studies.^{182–185} Cost effectiveness, possibility to investigate rare diseases, large cohorts allowing stratification, and analysis of co-occurring conditions are some of the advantages over classical RCTs.¹⁸⁰

Electronic health records in pharmacogenomics

Increasingly, biobanks match genotype entries with EHRs, broadening the phenotype query space exploitable for PGx research.¹⁸⁶ Table 2 lists major research initiatives and biobanks for which EHRs are available. The UK Biobank (UKB) recently added primary care data for ~230,000 participants, including consultation reports, medical diagnoses, and medication prescription.²⁰⁵ FinnGen plans to recruit ~500,000 individuals by 2023 with access to EHRs from national health registries comprehending medication usage.²⁰³ Similarly, the Estonian Biobank contains detailed information about drug purchase and disease incidences collected since 2000.^{183,187} In parallel, many initiatives focus on the clinical implementation of PGx,⁸ such as the American eMERGE-PGx project,¹²⁸ the European Ubiquitous Pharmacogenomics (U-PGx) consortium,²⁰⁶ the South East Asian Pharmacogenomics Research Network (SEAPharm),^{207,208} or the African Pharmacogenomics Consortium (APC).²⁰⁹ Because of the increased genetic diversity observed in African populations and their exposure to different types of diseases (e.g., higher infectious disease burden), the latter initiative is promising to unravel new PGx findings.

EHR-coupled biobanks accelerate our understanding of PGx by enabling the replication of known interactions and catalyzing new discoveries. Associations between drug maintenance dose and 9 PGx genes were tested in 200,000 UKB participants by assigning individuals to a metabolizer class (e.g., poor, intermediate, and normal) based on their genotype, revealing known *CYP2C9* and a novel *CYP2C19* variant as determinants for warfarin dosage.¹⁸⁴ Another UKB study discovered an association between statin use and *NAT2* (MIM: 612182) genotype.²¹⁰ The first study making use of WGS-coupled EHRs was conducted in ~2,500 Estonian Biobank participants.¹⁸³ Increasing sample size to ~16,000 by including individuals with imputed genotypes, six known and nine new PGx interactions were identified.¹⁸³ Importantly, many biobanks follow participants longitudinally, so that the number of individuals exposed to a drug or reporting an ADR increases with

time. Initiatives such as the International HundredK+ Cohorts Consortium (IHCC) aim at creating a platform to harmonize data and facilitate information sharing across cohorts.²¹¹ Specifically, one of IHCC's core initiatives is to "improve understanding of variability in response to treatments and identify novel PGx associations." With ~50 million participants from >100 cohorts,²¹¹ statistical power largely surpasses the one of traditional RCTs, making it likely for combined biobank studies to establish new PGx interactions.

Leveraging observational data to gain causal insights

Biobank studies follow the design of nonrandomized trials. Despite high agreement between treatment effects in randomized and nonrandomized trials,²¹² correlations between drug responses and genetic variants do not guarantee causality. Propensity score matching reduces bias from concomitant confounding variables through inclusion of selected covariates,¹⁸⁰ and randomization between study participants is mimicked based on known observed covariates. Alternatively, Mendelian randomization (MR) approaches account for unobserved confounders,²¹³ where the causal effect of a modifiable risk factor on an outcome is inferred from genetic associations coming from observational data (i.e., GWAS summary statistics or individual-level data). Analogously to RCTs, individuals are stratified based on the presence of genetic variants affecting the exposure of interest. MR was recently extended to study PGx through the triangulation within a study (TWIST) framework,²¹⁴ which estimates a genetically moderated treatment effect (GMTE) and allows to calculate the reduction in treatment effect experienced by patients with a particular PGx variant, as compared with individuals lacking the variant (Figure 5A). Applying TWIST to the UKB revealed that *CYP2C19* LOF carriers experienced a 0.28% increased risk of stroke ($p = 0.048$) while on clopidogrel.²¹⁴ As the TAILOR-PCI RCT failed at detecting a significant benefit to *CYP2C19*-based anti-platelet prescription (Randomized controlled trials), increased power resulting from larger sample size (clopidogrel $n = 7,483$; no clopidogrel $n = 198,868$) and consideration of a larger spectrum of LOF alleles could have contributed to the positive results. A different study analyzing statin usage in the UKB found that females homozygous for the *SLCO1B1*5* allele known to impair intracellular statin uptake²¹⁵ would have 0.147 mmol/L lower total cholesterol were they treated with a lipid-lowering medication unaffected by their genotype,²¹⁶ reiterating the importance of sex as a modulator of PGx interactions. Overall, MR studies provide *in silico* orthogonal lines of evidence for the functional consequence of PGx variants and represent valuable tools to prioritize RCTs.

Challenges of large-scale biomedical databases

Increased usage of EHR-linked biobanks has brought attention to new challenges. First, EHRs frequently consist of unstructured, noisy, sparse, and incomplete clinical notes. As

Table 2. List of 10 major biobanks and initiatives suitable for pharmacogenetic research

	Recruitment	Sample size	Sequencing data	EHR data	Omics data	References	IHCC	Notes
Estonian Genome Project	2002–	200,000	genome-wide genotyping (n = 200,000); WES (n = 2,500); WGS (n = 3,000)	n > 40,000	transcriptomics (n = 600; blood); metabolomics (n = 11,000); methylomics (n = 800); microbiomics (n = 2,500)	Leitsalu et al. ¹⁸⁷ , Leitsalu et al. ¹⁸⁸ , https://genomics.ut.ee	yes	in 2018–2019, increased from 50,000 to 200,000 samples
eMERGE	2007–	>135,000	genome-wide genotyping (n > 105,000); WES (n > 3,700); WGS (n > 1,700); PGRNseq (n = 9,000 ^a)	all participants are linked to EHRs	None	Gottesman et al. ¹⁸⁹ , Bush et al. ¹²⁹ , https://emerge-network.org	yes	network of multiple distinct cohorts
UK Biobank	2006–2010	500,000	genome-wide genotyping (n = 500,000); WES (n = 500,000); WGS (n = 500,000 ^{b,c})	n = 230,000	blood biomarkers (n = 500,000); proteomics (n = 53,000 ^b); metabolomics (n = 120,000); telomere length (n = 500,000)	Elliott and Peakman ¹⁹⁰ , Bycroft et al. ¹⁹¹ , Backman et al. ¹⁹² , https://www.ukbiobank.ac.uk	yes	
DiscovEHR	2007–	>250,000	genome-wide genotyping (n > 150,000); WES (n > 100,000)	all participants are linked to EHRs	blood biomarkers	Carey et al. ¹⁹³ , Dewey et al. ¹⁹⁴ , http://www.discovehrshare.com	yes	part of Geisinger's MyCode Community Health Initiative
Million Veteran Program	2011–	>1,000,000 ^b	genome-wide genotyping (n = 1,000,000 ^{b,d}); WES ^b ; WGS (n = 100,000)	all participants are linked to Veteran Affairs EHR	blood biomarkers (n = 1,000,000 ^b); proteomics ^b ; methylomics ^b	Gaziano et al. ¹⁹⁵ , Hunter-Zinck et al. ¹⁹⁶ , https://www.mvp.va.gov	yes	
Taiwan Biobank	2012–	200,000 ^e	genome-wide genotyping (n = 200,000 ^{e,f}); WGS (n = 1,500)	national EHRs available ^g	blood biomarkers; metabolomics ^e	Wei et al. ¹⁹⁷ , Lin et al. ¹⁹⁸ , https://www.twbiobank.org.tw	yes	
H3Africa	2012–	>100,000	genome-wide genotyping ^h ; WES ^h ; WGS ^h (n > 400)	None	blood biomarkers; microbiomics ^h	H3Africa Consortium et al. ¹⁹⁹ , Mulder et al. ²⁰⁰ , https://h3africa.org	no	H3Africa is composed of several cohorts
Tohoku Medical Megabank Project	2013–2017	150,000	genome-wide genotyping (n = 150,000 ^b); WGS (n = 14,000)	EHRs from MMWIN back-up system	transcriptomics (n = 100; blood); proteomics (n = 500); metabolomics (n = 46,000); methylomics (N = 100)	Tadaka et al. ²⁰¹ , Ido et al. ²⁰² , https://www.megabank.tohoku.ac.jp	no	composed of a population community-based cohort and a Birth and Three-Generation Cohort
FinnGen	2017–2023	500,000 ⁱ	genome-wide genotyping (n = 500,000 ^{i,j})	all participants are linked to national EHRs	None	Locke et al. ²⁰³ , https://www.finnngen.fi	no	
All of Us	2018–	>1,000,000 ^b	genomic assays ^b	all participants are linked to EHRs	biological assays ^b	The All of Us Research Program Investigators ²⁰⁴ , https://www.researchallofus.org	yes	

The IHCC column informs whether the cohort is part of the International HundredK + Cohorts Consortium.

^aeMERGE-PGx Project

^bPlanned.

^c150,000 samples released in 2021.

^d460,000 samples released in 2020.

^ePlanned in the population community-based cohort.

^f100,000 samples released in 2021.

^gNot yet crosslinked.

^hSample sizes and available measurements are cohort dependent.

ⁱBy 2023.

^j350,000 samples released in 2022.

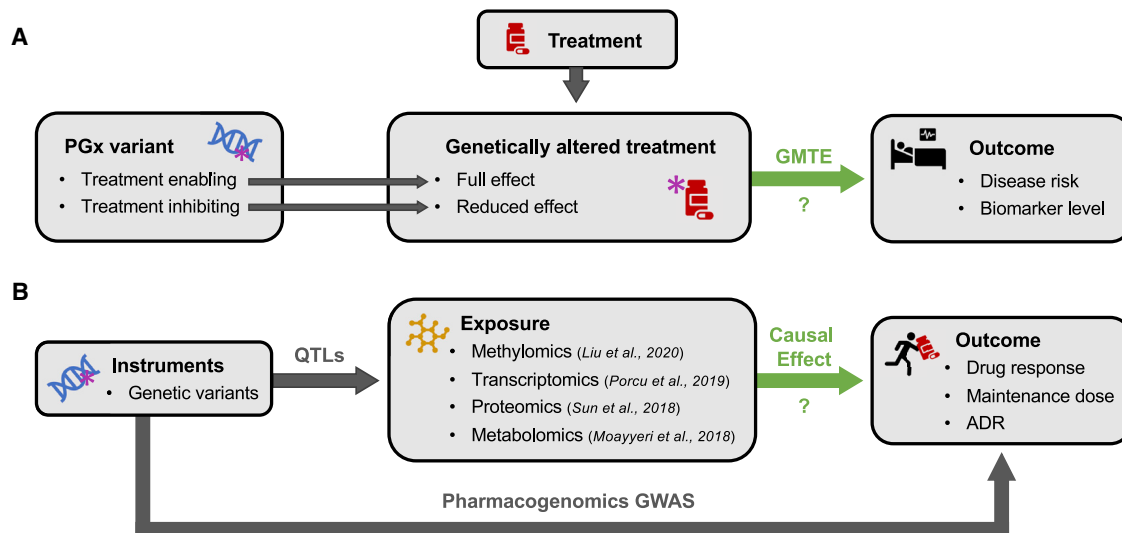


Figure 5. MR-analogous frameworks to infer causal effects in pharmacogenetic studies

(A) Framework proposed by Bowden et al.²¹⁴ to estimate a genetically modified treatment effect (GMTE) (green arrow), which describes the reduction in treatment effect experienced by patients with a PGx variant (treatment inhibiting) as opposed to those without (treatment enabling) the variant whilst on the same treatment.

(B) Proposed framework to identify causal molecular mediators of PGx effects. Genetic variants are used as instrumental variables, intermediate phenotypes (e.g., methylomics, transcriptomics, proteomics, or metabolomics data) as exposure, and drug responses as outcome. The causal effect of the exposure on the outcome (green arrow) is estimated thanks to (1) the effect of the genetic variants on the exposure, measured by means of quantitative trait loci (QTLs), and (2) the effect of the same genetic variants on the outcome, measured by means of PGx GWAS.

manual review by clinical experts is not scalable, automated solutions based on natural language-processing algorithms are required to extract informative features.²¹⁷ Second, the massive size of biobanks necessitates sufficient storage capacities and computational resources to conduct analyses.²¹⁸ This issue is exacerbated by the sensitive nature of the data: whereas patient privacy should be ensured, access to the data should be facilitated for the research community. With varying legislation across countries and differing views on acceptable privacy levels, development of data storage and management solutions is ongoing work.²¹⁹ Finally, validating the increasing number of putative PGx interactions resulting from data-driven approaches is likely to become a major bottleneck. Conventional methods, such as RCTs, are inherently slow and low throughput, calling for alternative ways to validate putative PGx findings.

Integration of multi-omics data

Biotechnological advances foster the generation of “omics” datasets, including methylomics, transcriptomics, proteomics, and metabolomics, which quantify DNA methylation, RNA expression, protein expression and modifications, and metabolite abundance, respectively.²²⁰ As disruptions in intermediate layers might be predictive of therapeutic outcomes (i.e., drug responses), exploring and integrating changes in biomarkers could help understand and predict pharmaco-omics interactions.

Molecular modeling of pharmaco-omics interactions

GWASs have shed light on the genetic architecture of complex traits but fall short on elucidating molecular

mechanisms,²²¹ as they identify associations between traits and blocks of co-inherited variants in LD, hindering fine-mapping. This challenge is addressed by methods integrating GWAS results with other omics data. Colocalization approaches prioritize putative causal genes based on shared signals between GWAS and omics data.^{222,223} Similarly, transcriptome-wide association studies (TWASs) pinpoint genes whose differential expression is associated with a given phenotype.^{224,225} Starting from genetic variants, gene expression is inferred from expression quantitative trait loci (eQTLs) and a derived gene expression PGS is correlated to the investigated trait. MR provides a framework that can be generalized to any modifiable exposure, including the transcriptome,²²⁶ proteome,^{227,228} methylome (CpG sites),²²⁹ or metabolome.²³⁰ Not widely employed in PGx,²³¹ these methods improve our understanding of the molecular basis of association signals and could provide a corroborating source of evidence for novel PGx interactions (Figure 5B).

Recent studies highlighted the benefits of including biomarkers to machine-learning (ML) algorithms predicting drug response²³² for antidepressant use,²³³ precision psychiatry,²³⁴ and warfarin dosage²⁷ (Box 2). In the latter study, assuming that warfarin dosage depends on genetic variation altering both the expression and activity of key enzymes (Figure 3), tissue-specific eQTLs were used to infer the expression of 116 genes implicated in warfarin pharmacokinetics and pharmacodynamics and construct prediction models, which improved predictive performance by an additional 8%–12% of variance.

Limitations to multi-omics modeling in pharmacogenomics

Despite promising results, most predictive models are not ready for the clinics. Limited size and number of available multi-omics datasets increase the risk of model overfitting and prevents independent replication, respectively.²³⁵ Furthermore, study results are often siloed and only few databases with unbiased drug-screening experiments are publicly available.²³⁶ There is a need for large-scale *in vitro* and *in vivo* perturbation experiments measuring drug effects on multiple omics layers,²³⁷ as results from these assays could open new avenues to study biomarker combinations and drug-drug-gene interactions.¹⁶¹ Importantly, omics layers are not confined entities. Instead, each forms a small portion of an inter-connected biological system, so that testing every possible drug-biomarker combination in isolation would not provide a complete picture. System genetics strategies have been proposed to elucidate molecular mechanisms,^{221,238} but they remain exploratory and often lack the drug-omics interaction dimension. Toward this direction, a “multiscale interactome” was constructed, compiling public data on interactions between drugs, diseases, proteins, and biological functions to identify disease treatment mechanisms.²³⁹ System-level PGx is likely to become more elaborate in the future ([Perspectives in mechanistic approaches](#)), providing mechanistic explanations for pharmaco-omics interactions.

In silico pharmacogenomics

Functional prediction of rare variants

The wealth of rare variants in known PGx genes discovered by NGS makes it challenging to infer their function through classical methods. Not only are *in silico* approaches necessary, but they can advance PGx faster and beyond what can be achieved with traditional methods. Functional impact of common variants can be characterized through association studies, *in vitro* heterologous expression systems, or computational prediction tools. However, these typically fall short when assessing rare variants.²⁴⁰ Computational algorithms usually rely on evolutionary conservation and use training sets with annotated pathogenic variants. Often under lower evolutionary constraints than disease-associated variants, PGx variants can be functional without being pathogenic.²⁴⁰ One solution is to engineer cell lines to harbor these variants and estimate their effect on enzymatic activity in a high-throughput and cost-effective manner, as was recently done for >8,000 *CYP2C9* variants¹⁷² following a forward genetics approach ([A forward genetics approach to pharmacogenetics](#)). Generated data can be used to train ML algorithms, yielding models applicable to clinical data. For instance, an ensemble ML model combining multiple functionality prediction algorithms was trained based on the enzymatic activity of 337 deleterious variants across 43 ADME genes, significantly increasing performance accuracy compared with single ML algorithms.²⁴¹

Variant combinations—within the same or different genes—can also affect drug response. Employing deep learning algorithms, haplotype effects were investigated to predict *CYP2D6* enzymatic activity.^{242,243} Training a convolutional neural network based on star alleles definitions, *in vitro* enzymatic activity of both known and unknown *CYP2D6* haplotypes could be predicted.²⁴² Similarly, a neural network was trained on a prospective cohort of breast cancer patients to predict *CYP2D6*-mediated tamoxifen metabolism.²⁴³ In this second study, omitting star allele haplotype assignment yielded better prediction scores, implying it might be time to move beyond star allele PGx. Both studies being limited in their scope, investigating a broader set of genes is warranted to draw general conclusions.

Perspectives in mechanistic approaches

We previously introduced statistical methods that integrate GWAS findings with other datasets to identify causal biomarker-drug response relationships. Assuming a biomarker alteration is causal to a disease, identifying drugs achieving their therapeutic effect by inducing the opposite change should provide viable treatment options. This hypothesis was tested by searching for drugs inducing gene expression patterns opposite to the ones observed in seven psychiatric disorders within the Connectivity Map (CMap),²⁴⁴ a library containing the expression profile of 1,000 genes upon perturbation by ~20,000 small molecules.²⁴⁵ Although there is no protein analogue to CMap, ChEMBL, the largest database of curated drug targets,²⁴⁶ and DrugBank⁷⁸ form valuable resources to conduct similar studies at the proteome level. Drug repositioning is a topic on its own,²⁴⁷ but this example showcases an application of a mechanistic understanding of drug-disease interactions. In the context of PGx, this strategy could identify drugs counteracting patient-specific, disease-causing biomarker changes.

Alternatively to data-driven top-down strategies are bottom-up approaches, such as the ones used in systems biology.²⁴⁸ Systems biology aims at constructing metabolic network models describing complex biological systems,²⁴⁹ with nodes representing metabolite concentrations connected by mathematical equations defined by enzyme kinetics. Since enzyme kinetics is at the core of drug metabolism,²⁵⁰ it would be relevant to understand how consequences of genetic variants propagate through the biochemical network ([Figure 6](#)). Network effects can be calculated through flux balance analyses, using mathematical models accounting for thermodynamic constraints and mRNA and protein abundances.²⁵¹ Efforts to reconstruct a genome-scale network are concerted in Recon3D, the most comprehensive human metabolic network to date.²⁵² Recon3D provides information on 13,543 enzymatic reactions and 4,140 metabolites and describes the relationships between involved genes, proteins, the reactions these proteins catalyze, and links between genetic variants and metabolic effects. Integration of genetic

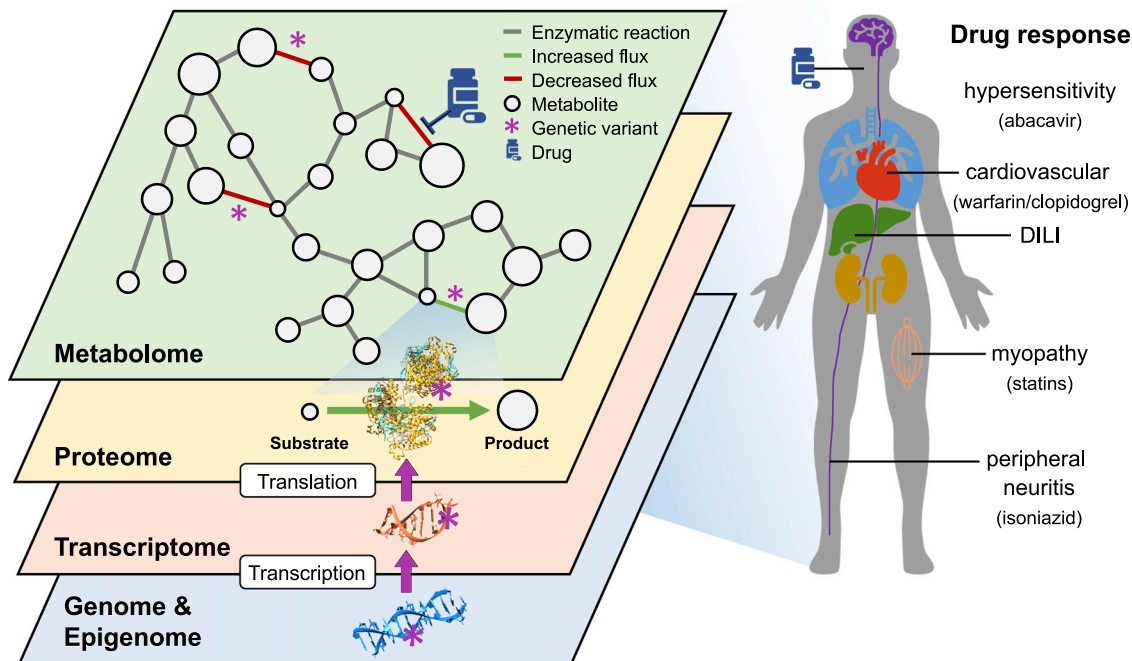


Figure 6. A mechanistic systems biology view on pharmacogenomics

The metabolic network of each cell is composed of metabolites (nodes) connected by enzymatic reactions (edges). It is modulated by genetic and epigenetic variants (star) that propagate through several intermediate molecular layers (e.g., transcriptome and proteome), as well as by drug intake. This network and its modulation are cell type specific and influence an individual's drug response and ADR risk, as illustrated on the right with examples of differential drug responses manifesting themselves in specific tissues.

variation into metabolic studies remains in its infancy, but the ever-growing supply of diverse datasets is leading to a steady increase in our understanding of disease mechanisms. There is a great demand for refined multiscale modeling, as expressed in a review authored by the pharmaceutical industry, with the hope that more realistic *in silico* models will improve prediction of drug mechanisms and safety before human clinical testing.²⁵³

The past and future of pharmacogenetics

Seventy years of methodological development have transformed PGx from an emerging science into an interdisciplinary area of research key to the implementation of personalized medicine. Pharmacogenetics established the role of genetics in pharmacological responses, with RCTs validating the first CGSs discoveries. Completion of the Human Genome Project triggered the transition to pharmacogenomics: sequencing technologies allowed ascertainment of the human PGx landscape while genome-editing tools enabled discovery and functional validation of PGx interactions. Finally, tools and concepts from statistics and computer science have been and will be key to analyze and integrate ever more complex datasets, bringing us a step closer to the *in silico* study of pharmaco-omics.

Despite indisputable progress, numerous challenges lay ahead. Throughout the review, we highlighted areas of research we imagine will be decisive for the future of PGx

and that overlap challenges recently laid out.²⁵⁴ Due to its interdisciplinary nature, progress in PGx can be achieved through different avenues, including data generation, data analysis and integration, and discovery prioritization and validation. Size and diversity of information contained in biobanks is constantly increasing, as development of sequencing and omics technologies improves genetic and molecular pathway understanding, while EHRs broaden clinical characterization. In parallel, high-throughput screens in genetically engineered cell lines, patient-derived organoids, and model organisms can assess the impact of genetic variations in controlled but tunable environmental settings. Establishment of efficient and secure solutions for data storage and sharing will foster the development of statistical and ML approaches aiming at analyzing and integrating data originating from various sources, resulting in a better understanding of the genetic architecture of PGx traits and mechanistic insights. RCTs will likely continue to play an important role in the clinical implementation of PGx, but it will be important to maximize their success rate and improve their throughput. Creative *in vitro* and *in vivo* experiments, as well as high-quality *in silico* mechanistic and predictive models, can generate robust hypotheses, thereby improving prioritization of new putative PGx interactions. Awareness and incorporation of this knowledge in RCT design might increase their success rate and/or promote development of companion diagnostics gauging the suitability of a treatment for a given patient.²⁵⁵

Solutions to the above-mentioned challenges should ensure that progress in PGx benefits all. Genomic research remains predominantly conducted in populations of European ancestry, despite general agreement that more diversity is needed to understand the human genome.^{24,256–259} This particularly applies to PGx, as many PGx genes are under weak global selective pressures and exhibit population-specific allele frequencies.^{15–19,130,260} In this optic, *All of US* will recruit one million US citizens to build the largest and most diverse cohort in history, marking an important shift from the paradigm of studying white-male individuals to representing the full spectrum of a globalized world, including its minorities.²⁰⁴ Another core value of the project is to strengthen ties between participants and researchers to return value to all parties and ensure knowledge flow from bench to bedside and back. Acknowledging that most health care systems currently lack the infrastructure to implement screening for even well-established PGx variant, we argue that genetically-guided prescription, together with the recording of health outcomes through EHRs, will fuel PGx research while helping refine prescription guidelines. This virtuous cycle will benefit from increased diversity in both research and clinical settings, as the richness in genetic variation will accelerate mechanistic understandings of PGx, whose findings can in turn be generalized to other populations.

Acknowledgments

This work was supported by funding from the Department of Computational Biology (Z.K.) and the Center for Integrative Genomics (A.R.) from the University of Lausanne, as well as funding from the Swiss National Science Foundation (310030-189147 to Z.K.). We thank Prof. Christophe Dessimoz for sparking the idea of this review and Prof. Johan Auwerx for discussions and critiques. Due to the large amount of covered material, we had to carefully select references and apologize for any omissions.

Declaration of interests

The authors declare no competing interests.

Web resources

ClinicalTrials.gov, <https://clinicaltrials.gov/>.
 CPIC, <https://cpicpgx.org/>.
 IHCC, <https://ihccglobal.org/>.
 OMIM, <https://www.omim.org/>.
 PharmVar, <https://www.pharmvar.org/>.
 PubMed.gov, <https://pubmed.ncbi.nlm.nih.gov/>.

References

1. Relling, M.V., and Evans, W.E. (2015). Pharmacogenomics in the clinic. *Nature* 526, 343–350.
2. Lauschke, V.M., Zhou, Y., and Ingelman-Sundberg, M. (2019). Novel genetic and epigenetic factors of importance for inter-individual differences in drug disposition, response and toxicity. *Pharmacol. Ther.* 197, 122–152.

3. Relling, M.V., Klein, T.E., Gammal, R.S., Whirl-Carrillo, M., Hoffman, J.M., and Caudle, K.E. (2020). The clinical pharmacogenetics implementation consortium: 10 years later. *Clin. Pharmacol. Ther.* 107, 171–175.
4. Dunnenberger, H.M., Crews, K.R., Hoffman, J.M., Caudle, K.E., Broeckel, U., Howard, S.C., Hunkler, R.J., Klein, T.E., Evans, W.E., and Relling, M.V. (2015). Preemptive clinical pharmacogenetics implementation: current programs in five US medical centers. *Annu. Rev. Pharmacol. Toxicol.* 55, 89–106.
5. Alshabeeb, M.A., Deneer, V.H., and Asselbergs, F.W. (2019). Use of pharmacogenetic drugs by the Dutch population. *Front. Genet.* 10, 567.
6. Schildcrout, J.S., Denny, J.C., Bowton, E., Gregg, W., Pulley, J.M., Basford, M.A., Cowan, J.D., Xu, H., Ramirez, A.H., Crawford, D.C., et al. (2012). Optimizing drug outcomes through pharmacogenetics: a case for preemptive genotyping. *Clin. Pharmacol. Ther.* 92, 235–242.
7. Chanfreau-Coffinier, C., Hull, L.E., Lynch, J.A., DuVall, S.L., Damrauer, S.M., Cunningham, F.E., Voight, B.F., Matheny, M.E., Oslin, D.W., Icardi, M.S., et al. (2019). Projected prevalence of actionable pharmacogenetic variants and level a drugs prescribed among US veterans health administration pharmacy users. *JAMA Netw. Open* 2, e195345.
8. Krebs, K., and Milani, L. (2019). Translating pharmacogenomics into clinical decisions: do not let the perfect be the enemy of the good. *Hum. Genomics* 13, 1–13.
9. Klein, M.E., Parvez, M.M., and Shin, J.-G. (2017). Clinical implementation of pharmacogenomics for personalized precision medicine: barriers and solutions. *J. Pharm. Sci.* 106, 2368–2379.
10. Vogel, F. (1959). Moderne probleme der Humangenetik. In *Ergebnisse der Inneren Medizin und Kinderheilkunde* (Springer), pp. 52–125.
11. Haldane, J.B.S. (1954). *The Biochemistry of Genetics* (George Allen & Unwin).
12. Motulsky, A.G. (1957). Drug reactions, enzymes, and biochemical genetics. *JAMA* 165, 835–837.
13. Hockwald, R.S., Arnold, M.J., Clayman, C.B., and Alving, A.S. (1952). Toxicity of primaquine in Negroes. *JAMA* 149, 1568–1570.
14. Nkhoma, E.T., Poole, C., Vannappagari, V., Hall, S.A., and Beutler, E. (2009). The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. *Blood Cell Mol. Dis.* 42, 267–278.
15. Li, J., Zhang, L., Zhou, H., Stoneking, M., and Tang, K. (2011). Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum. Mol. Genet.* 20, 528–540.
16. Ramos, E., Doumatey, A., Elkahloun, A., Shriner, D., Huang, H., Chen, G., Zhou, J., McLeod, H., Adeyemo, A., and Rotimi, C. (2014). Pharmacogenomics, ancestry and clinical decision making for global populations. *Pharmacogenomics J.* 14, 217–222.
17. Hovelson, D.H., Xue, Z., Zawistowski, M., Ehm, M.G., Harris, E.C., Stocker, S.L., Gross, A.S., Jang, I.-J., Ieiri, I., Lee, J.-E., et al. (2017). Characterization of ADME gene variation in 21 populations by exome sequencing. *Pharmacogenetics Genom.* 27, 89–100.
18. Zhou, Y., Ingelman-Sundberg, M., and Lauschke, V.M. (2017). Worldwide distribution of cytochrome P450 alleles: a meta-analysis of population-scale sequencing projects. *Clin. Pharmacol. Ther.* 102, 688–700.

19. Ahn, E., and Park, T. (2017). Analysis of population-specific pharmacogenomic variants using next-generation sequencing data. *Sci. Rep.* *7*, 1–11.
20. McInnes, G., Yee, S.W., Pershad, Y., and Altman, R.B. (2021). Genomewide association studies in pharmacogenomics. *Clin. Pharmacol. Ther.* *110*, 637–648.
21. McInnes, G.M., Lavertu, A., Sangkuhl, K., Klein, T.E., Whirl-Carrillo, M., and Altman, R.B. (2021). Pharmacogenetics at scale: an analysis of the UK Biobank. *Clin. Pharmacol. Ther.* *109*, 1528–1537.
22. FDA (2020). 2019 Drug Trials Snapshots Summary Report, Table 1. Demographic Subgroups in 2019. <https://www.fda.gov/media/135337/download>.
23. Huddart, R., Sangkuhl, K., Whirl-Carrillo, M., and Klein, T.E. (2019). Are randomized controlled trials necessary to establish the value of implementing pharmacogenomics in the clinic? *Clin. Pharmacol. Ther.* *106*, 284–286.
24. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* *51*, 584–591.
25. Ge, D., Fellay, J., Thompson, A.J., Simon, J.S., Shianna, K.V., Urban, T.J., Heinzen, E.L., Qiu, P., Bertelsen, A.H., Muir, A.J., et al. (2009). Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* *461*, 399–401.
26. Mak, A.C., White, M.J., Eckalbar, W.L., Szpiech, Z.A., Oh, S.S., Pino-Yanes, M., Hu, D., Goddard, P., Huntsman, S., Galanter, J., et al. (2018). Whole-genome sequencing of pharmacogenetic drug response in racially diverse children with asthma. *Am. J. Respir. Crit. Care Med.* *197*, 1552–1564.
27. Gottlieb, A., Daneshjou, R., DeGorter, M., Bourgeois, S., Svensson, P.J., Wadelius, M., Deloukas, P., Montgomery, S.B., and Altman, R.B. (2017). Cohort-specific imputation of gene expression improves prediction of warfarin dose for African Americans. *Genome Med.* *9*, 98.
28. Vesell, E.S. (1989). Pharmacogenetic perspectives gained from twin and family studies. *Pharmacol. Ther.* *41*, 535–552.
29. Mahgoub, A., Dring, L., Idle, J., Lancaster, R., and Smith, R. (1977). Polymorphic hydroxylation of Debrisoquine in man. *Lancet* *310*, 584–586.
30. Eichelbaum, M., Spannbrucker, N., Steincke, B., and Dengler, H. (1979). Defective N-oxidation of sparteine in man: a new pharmacogenetic defect. *Eur. J. Clin. Pharmacol.* *16*, 183–187.
31. Inaba, T., Otton, S., and Kalow, W. (1980). Deficient metabolism of debrisoquine and sparteine. *Clin. Pharmacol. Ther.* *27*, 547–549.
32. Eichelbaum, M., Bertilsson, L., Säwe, J., and Zekorn, C. (1982). Polymorphic oxidation of sparteine and debrisoquine: related pharmacogenetic entities. *Clin. Pharmacol. Ther.* *31*, 184–186.
33. Peters, B.J., Rodin, A.S., De Boer, A., and Maitland-van der Zee, A.-H. (2010). Methodological and statistical issues in pharmacogenomics. *J. Pharm. Pharmacol.* *62*, 161–166.
34. Distlerath, L.M., Reilly, P., Martin, M., Davis, G., Wilkinson, G., and Guengerich, F. (1985). Purification and characterization of the human liver cytochromes P-450 involved in debrisoquine 4-hydroxylation and phenacetin O-deethylation, two prototypes for genetic polymorphism in oxidative drug metabolism. *J. Biol. Chem.* *260*, 9057–9067.
35. Gonzalez, F.J., Skodat, R.C., Kimura, S., Umeno, M., Zanger, U.M., Nebert, D.W., Gelboin, H.V., Hardwick, J.P., and Meyer, U.A. (1988). Characterization of the common genetic defect in humans deficient in debrisoquine metabolism. *Nature* *331*, 442–446.
36. Zanger, U.M., and Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* *138*, 103–141.
37. Hetherington, S., McGuirk, S., Powell, G., Cutrell, A., Naderer, O., Spreen, B., Lafon, S., Pearce, G., and Steel, H. (2001). Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. *Clin. Ther.* *23*, 1603–1614.
38. Hetherington, S., Hughes, A.R., Mosteller, M., Shortino, D., Baker, K.L., Spreen, W., Lai, E., Davies, K., Handley, A., Dow, D.J., et al. (2002). Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. *Lancet* *359*, 1121–1122.
39. Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A., Moore, C., Sayer, D., Castley, A., Mamotte, C., Maxwell, D., et al. (2002). Association between presence of HLA-B* 5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* *359*, 727–732.
40. NHGRI (2020). The Cost of Sequencing a Human Genome. <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>.
41. Ross, S., Anand, S.S., Joseph, P., and Paré, G. (2012). Promises and challenges of pharmacogenetics: an overview of study design, methodological and statistical issues. *JRSM Cardiovasc. Dis.* *1*, 1–13.
42. Meldrum, M.L. (2000). A brief history of the randomized controlled trial: from oranges and lemons to the gold standard. *Hematology/oncol. Clin.* *14*, 745–760.
43. Mallal, S., Phillips, E., Carosi, G., Molina, J.-M., Workman, C., Tomažič, J., Jägel-Guedes, E., Rugina, S., Kozyrev, O., Cid, J.F., et al. (2008). HLA-B* 5701 screening for hypersensitivity to abacavir. *N. Engl. J. Med.* *358*, 568–579.
44. Department of Health and Human Services (2021). Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV. <https://clinicalinfo.hiv.gov/sites/default/files/guidelines/documents/AdultandAdolescentGL.pdf>.
45. Mounzer, K., Hsu, R., Fusco, J.S., Brunet, L., Henegar, C.E., Vannappagari, V., Stainsby, C.M., Shaefer, M.S., Ragone, L., and Fusco, G.P. (2019). HLA-B* 57:01 screening and hypersensitivity reaction to abacavir between 1999 and 2016 in the OPERA® observational database: a cohort study. *AIDS Res. Ther.* *16*, 1–9.
46. Pirmohamed, M., and Hughes, D.A. (2013). Pharmacogenetic tests: the need for a level playing field. *Nat. Rev. Drug Discov.* *12*, 3–4.
47. Moore, T.J., Zhang, H., Anderson, G., and Alexander, G.C. (2018). Estimated costs of pivotal trials for novel therapeutic agents approved by the US Food and Drug Administration, 2015–2016. *JAMA Intern. Med.* *178*, 1451–1457.
48. Pereira, N.L., Farkouh, M.E., So, D., Lennon, R., Geller, N., Mathew, V., Bell, M., Bae, J.-H., Jeong, M.H., Chavez, I., et al. (2020). Effect of genotype-guided oral P2Y12 inhibitor selection vs conventional clopidogrel therapy on ischemic outcomes after percutaneous coronary intervention: the TAILOR-PCI randomized clinical trial. *JAMA* *324*, 761–771.

49. Pereira, N.L., Rihal, C.S., So, D.Y., Rosenberg, Y., Lennon, R.J., Mathew, V., Goodman, S.G., Weinsilboum, R.M., Wang, L., Baudhuin, L.M., et al. (2019). Clopidogrel pharmacogenetics: state-of-the-art review and the TAILOR-PCI study. *Circ. Cardiovasc. Interventions* 12, e007811.
50. Mega, J.L., Simon, T., Collet, J.-P., Anderson, J.L., Antman, E.M., Bliden, K., Cannon, C.P., Danchin, N., Giusti, B., Gurbel, P., et al. (2010). Reduced-function CYP2C19 genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. *JAMA* 304, 1821–1830.
51. Kazui, M., Nishiya, Y., Ishizuka, T., Hagihara, K., Farid, N.A., Okazaki, O., Ikeda, T., and Kurihara, A. (2010). Identification of the human cytochrome P450 enzymes involved in the two oxidative steps in the bioactivation of clopidogrel to its pharmacologically active metabolite. *Drug Metab. Dispos.* 38, 92–99.
52. Lewis, J.P., Backman, J.D., Reny, J.-L., Bergmeijer, T.O., Mitchell, B.D., Ritchie, M.D., Déry, J.-P., Pakyz, R.E., Gong, L., Ryan, K., et al. (2020). Pharmacogenomic polygenic response score predicts ischaemic events and cardiovascular mortality in clopidogrel-treated patients. *Eur. Heart J. - Cardiovasc. Pharmacother.* 6, 203–210.
53. Galli, M., Benenati, S., Capodanno, D., Franchi, F., Rollini, F., D'Amario, D., Porto, I., and Angiolillo, D.J. (2021). Guided versus standard antiplatelet therapy in patients undergoing percutaneous coronary intervention: a systematic review and meta-analysis. *Lancet* 397, 1470–1483.
54. Pereira, N.L., Rihal, C., Lennon, R., Marcus, G., Shrivastava, S., Bell, M.R., So, D., Geller, N., Goodman, S.G., Hasan, A., et al. (2021). Effect of CYP2C19 genotype on ischemic outcomes during oral P2Y12 inhibitor therapy: a meta-analysis. *JACC: Cardiovasc. Interventions* 14, 739–750.
55. Lee, C.R., Luzum, J.A., Sangkuhl, K., Gammal, R.S., Sabatine, M.S., Stein, C.M., Kisor, D.F., Limdi, N.A., Lee, Y.M., Scott, S.A., et al. (2022). Clinical Pharmacogenetics Implementation Consortium guideline for CYP2C19 genotype and clopidogrel therapy: 2022 update. *Clin. Pharmacol. Ther.* 0, 1–9.
56. Pereira, N.L., Avram, R., So, D.Y., Iturriaga, E., Byrne, J., Lennon, R.J., Murthy, V., Geller, N., Goodman, S.G., Rihal, C., et al. (2021). Rationale and design of the TAILOR-PCI digital study: transitioning a randomized controlled trial to a digital registry. *Am. Heart J.* 232, 84–93.
57. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
58. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351.
59. Hopkins, A.L., and Groom, C.R. (2002). The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730.
60. Verlouw, J.A., Clemens, E., de Vries, J.H., Zolk, O., Verkerk, A.J., am Zehnhoff-Dinnesen, A., Medina-Gomez, C., Lanvers-Kaminsky, C., Rivadeneira, F., Langer, T., et al. (2021). A comparison of genotyping arrays. *Eur. J. Hum. Genet.* 29, 1611–1624.
61. Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 11, 499–511.
62. Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351.
63. Gilly, A., Southam, L., Suveges, D., Kuchenbaecker, K., Moore, R., Melloni, G.E., Hatzikotoulas, K., Farmaki, A.-E., Ritchie, G., Schwartzentruber, J., et al. (2019). Very low-depth whole-genome sequencing in complex trait association studies. *Bioinformatics* 35, 2555–2561.
64. Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., and Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* 53, 120–126.
65. Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Diltthey, A.T., Fiddes, I.T., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345.
66. Payne, A., Holmes, N., Rakyán, V., and Loose, M. (2019). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35, 2193–2198.
67. Pollard, M.O., Gurdasani, D., Mentzer, A.J., Porter, T., and Sandhu, M.S. (2018). Long reads: their purpose and place. *Hum. Mol. Genet.* 27, R234–R241.
68. Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.
69. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424.
70. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353, aaf8729.
71. Adikusuma, F., Williams, N., Grutzner, F., Hughes, J., and Thomas, P. (2017). Targeted deletion of an entire chromosome using CRISPR/Cas9. *Mol. Ther.* 25, 1736–1738.
72. Zuo, E., Huo, X., Yao, X., Hu, X., Sun, Y., Yin, J., He, B., Wang, X., Shi, L., Ping, J., et al. (2017). CRISPR/Cas9-mediated targeted chromosome elimination. *Genome Biol.* 18, 224.
73. Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* 9, 1911.
74. Pickar-Oliver, A., and Gersbach, C.A. (2019). The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* 20, 490–507.
75. Doudna, J.A. (2020). The promise and challenge of therapeutic genome editing. *Nature* 578, 229–236.
76. Relling, M., and Klein, T. (2011). CPIC: clinical pharmacogenetics implementation consortium of the pharmacogenomics research network. *Clin. Pharmacol. Ther.* 89, 464–467.
77. Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., Zhang, R., Zhu, J., Ren, Y., Tan, Y., et al. (2020). Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* 48, D1031–D1041.
78. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082.
79. Griffith, M., Griffith, O.L., Coffman, A.C., Weible, J.V., McMichael, J.F., Spies, N.C., Koval, J., Das, I., Callaway,

- M.B., Eldred, J.M., et al. (2013). DGIdb: mining the druggable genome. *Nat. Methods* *10*, 1209–1210.
80. Freshour, S.L., Kiwala, S., Cotto, K.C., Coffman, A.C., McMichael, J.F., Song, J.J., Griffith, M., Griffith, O.L., and Wagner, A.H. (2021). Integration of the drug–gene interaction database (DGIdb 4.0) with open crowdsourcing efforts. *Nucleic Acids Res.* *49*, D1144–D1151.
 81. Relling, M., Krauss, R., Roden, D., Klein, T., Fowler, D., Terada, N., Lin, L., Riel-Mehan, M., Do, T., Kubo, M., et al. (2017). New pharmacogenomics research network: an open community catalyzing research and translation in precision medicine. *Clin. Pharmacol. Ther.* *102*, 897–902.
 82. Whirl-Carrillo, M., McDonagh, E.M., Hebert, J., Gong, L., Sangkuhl, K., Thorn, C., Altman, R.B., and Klein, T.E. (2012). Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* *92*, 414–417.
 83. Whirl-Carrillo, M., Huddart, R., Gong, L., Sangkuhl, K., Thorn, C.F., Whaley, R., and Klein, T.E. (2021). An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* *110*, 563–572.
 84. Gaedigk, A., Ingelman-Sundberg, M., Miller, N.A., Leeder, J.S., Whirl-Carrillo, M., Klein, T.E., and Committee, P.S. (2018). The Pharmacogene Variation (PharmVar) consortium: incorporation of the human cytochrome P450 (CYP) allele nomenclature database. *Clin. Pharmacol. Ther.* *103*, 399–401.
 85. Gaedigk, A., Sangkuhl, K., Whirl-Carrillo, M., Twist, G.P., Klein, T.E., Miller, N.A.; and PharmVar Steering Committee (2019). The evolution of PharmVar. *Clin. Pharmacol. Ther.* *105*, 29–32.
 86. SEARCH Collaborative Group (2008). SLCO1B1 variants and statin-induced myopathy—a genomewide study. *N. Engl. J. Med.* *359*, 789–799.
 87. Hirano, M., Maeda, K., Shitara, Y., and Sugiyama, Y. (2004). Contribution of OATP2 (OATP1B1) and OATP8 (OATP1B3) to the hepatic uptake of pitavastatin in humans. *J. Pharmacol. Exp. Ther.* *311*, 139–146.
 88. Daly, A.K. (2010). Genome-wide association studies in pharmacogenomics. *Nat. Rev. Genet.* *11*, 241–246.
 89. Giacomini, K.M., Yee, S.W., Mushiroda, T., Weinshilboum, R.M., Ratain, M.J., and Kubo, M. (2017). Genome-wide association studies of drug response and toxicity: an opportunity for genome medicine. *Nat. Rev. Drug Discov.* *16*, 1–3.
 90. Yee, S.W., Momozawa, Y., Kamatani, Y., Tyndale, R.F., Weinshilboum, R.M., Ratain, M.J., Giacomini, K.M., and Kubo, M. (2016). Genomewide association studies in pharmacogenomics: meeting report of the NIH pharmacogenomics research network-RIKEN (PGRN-RIKEN) collaboration. *Clin. Pharmacol. Ther.* *100*, 423–426.
 91. Cooper, G.M., Johnson, J.A., Langae, T.Y., Feng, H., Stanaway, I.B., Schwarz, U.I., Ritchie, M.D., Stein, C.M., Roden, D.M., Smith, J.D., et al. (2008). A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* *112*, 1022–1027.
 92. Takeuchi, F., McGinnis, R., Bourgeois, S., Barnes, C., Eriksson, N., Soranzo, N., Whittaker, P., Ranganath, V., Kumanduri, V., McLaren, W., et al. (2009). A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet.* *5*, e1000433.
 93. Johnson, J.A., Gong, L., Whirl-Carrillo, M., Gage, B.F., Scott, S.A., Stein, C., Anderson, J., Kimmel, S.E., Lee, M.T.M., Pirmohamed, M., et al. (2011). Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clin. Pharmacol. Ther.* *90*, 625–629.
 94. Wadelius, M., and Pirmohamed, M. (2007). Pharmacogenetics of warfarin: current status and future challenges. *Pharmacogenomics J.* *7*, 99–111.
 95. Di Bisceglie, A.M., and Hoofnagle, J.H. (2002). Optimal therapy of hepatitis C. *Hepatology* *36*, S121–S127.
 96. Suppiah, V., Moldovan, M., Ahlenstiel, G., Berg, T., Weltman, M., Abate, M.L., Bassendine, M., Spengler, U., Dore, G.J., Powell, E., et al. (2009). IL28B is associated with response to chronic hepatitis C interferon- α and ribavirin therapy. *Nat. Genet.* *41*, 1100–1104.
 97. Tanaka, Y., Nishida, N., Sugiyama, M., Kurosaki, M., Matsuura, K., Sakamoto, N., Nakagawa, M., Korenaga, M., Hino, K., Hige, S., et al. (2009). Genome-wide association of IL28B with response to pegylated interferon- α and ribavirin therapy for chronic hepatitis C. *Nat. Genet.* *41*, 1105–1109.
 98. Rauch, A., Kutalik, Z., Descobes, P., Cai, T., Di Iulio, J., Mueller, T., Bochud, M., Battagay, M., Bernasconi, E., Borovicka, J., et al. (2010). Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* *138*, 1338–1345.
 99. Alexopoulou, A., and Karayiannis, P. (2015). Interferon-based combination treatment for chronic hepatitis C in the era of direct acting antivirals. *Ann. Gastroenterol.* *28*, 55–65.
 100. Falade-Nwulia, O., Suarez-Cuervo, C., Nelson, D.R., Fried, M.W., Segal, J.B., and Sulkowski, M.S. (2017). Oral direct-acting agent therapy for hepatitis C virus infection: a systematic review. *Ann. Intern. Med.* *166*, 637–648.
 101. Garrison, K.L., German, P., Mogalian, E., and Mathias, A. (2018). The drug-drug interaction potential of antiviral agents for the treatment of chronic hepatitis C infection. *Drug Metab. Dispos.* *46*, 1212–1225.
 102. Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* *52*, 399–433.
 103. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell* *169*, 1177–1186.
 104. Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans effects on gene expression can drive omnigenic inheritance. *Cell* *177*, 1022–1034.
 105. Shu, Y., Sheardown, S.A., Brown, C., Owen, R.P., Zhang, S., Castro, R.A., Ianculescu, A.G., Yue, L., Lo, J.C., Burchard, E.G., et al. (2007). Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J. Clin. Invest.* *117*, 1422–1431.
 106. Song, I., Shin, H., Shim, E., Jung, I., Kim, W., Shon, J., and Shin, J. (2008). Genetic variants of the organic cation transporter 2 influence the disposition of metformin. *Clin. Pharmacol. Ther.* *84*, 559–562.
 107. Becker, M.L., Visser, L.E., Van Schaik, R.H., Hofman, A., Uitterlinden, A.G., and Stricker, B.H.C. (2009). Genetic variation in the multidrug and toxin extrusion 1 transporter protein influences the glucose-lowering effect of metformin in patients with diabetes: a preliminary study. *Diabetes* *58*, 745–749.
 108. Jablonski, K.A., McAteer, J.B., de Bakker, P.I., Franks, P.W., Pollin, T.I., Hanson, R.L., Saxena, R., Fowler, S., Shuldiner,

- A.R., Knowler, W.C., et al. (2010). Common variants in 40 genes assessed for diabetes incidence and response to metformin and lifestyle intervention in the diabetes prevention program. *Diabetes* 59, 2672–2681.
109. Zhou, K., Donnelly, L., Yang, J., Li, M., Deshmukh, H., Van Zuydam, N., Ahlqvist, E., Spencer, C.C., Groop, L., Morris, A.D., et al. (2014). Heritability of variation in glycaemic response to metformin: a genome-wide complex trait analysis. *Lancet Diabetes Endocrinol.* 2, 481–487.
 110. Zhou, K., Bellenguez, C., Spencer, C.C., Bennett, A.J., Coleman, R.L., Tavendale, R., Hawley, S.A., Donnelly, L.A., Schofield, C., Groves, C.J., et al. (2011). Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nat. Genet.* 43, 117–120.
 111. Zhou, K., Yee, S.W., Seiser, E.L., Van Leeuwen, N., Tavendale, R., Bennett, A.J., Groves, C.J., Coleman, R.L., Van Der Heijden, A.A., Beulens, J.W., et al. (2016). Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. *Nat. Genet.* 48, 1055–1059.
 112. Rotroff, D.M., Yee, S.W., Zhou, K., Marvel, S.W., Shah, H.S., Jack, J.R., Havener, T.M., Hedderson, M.M., Kubo, M., Herman, M.A., et al. (2018). Genetic variants in CPA6 and PRPF31 are associated with variation in response to metformin in individuals with type 2 diabetes. *Diabetes* 67, 1428–1440.
 113. Zhang, Y., Hu, Y., Ho, K., Hartzel, D.N., Abedi, V., Zand, R., et al. (2020). Variants in STAU2 associate with metformin response in a type 2 diabetes cohort: a pharmacogenomics study using real-world electronic health record data. Preprint at MedRxiv. <https://doi.org/10.1101/2020.03.18.20037218>.
 114. Muhammad, A., Aka, I.T., Birdwell, K.A., Gordon, A.S., Roden, D.M., Wei, W.-Q., Mosley, J.D., and Van Driest, S.L. (2021). Genome-wide approach to measure variant-based heritability of drug outcome phenotypes. *Clin. Pharmacol. Ther.* 110, 714–722.
 115. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* 50, 1219–1224.
 116. Johnson, D., Wilke, M.A., Lyle, S.M., Kowalec, K., Jorgensen, A., Wright, G.E., and Drögemöller, B.I. (2021). A systematic review and analysis of the use of polygenic scores in pharmacogenomics. *Clin. Pharmacol. Ther.* <https://doi.org/10.1002/cpt.2520>.
 117. Thomas, L., Kessler, D., Campbell, J., Morrison, J., Peters, T.J., Williams, C., Lewis, G., and Wiles, N. (2013). Prevalence of treatment-resistant depression in primary care: cross-sectional data. *Br. J. Gen. Pract.* 63, e852–e858.
 118. Pain, O., Hodgson, K., Trubetskoy, V., Ripke, S., Marshe, V.S., Adams, M.J., et al. (2020). Antidepressant response in major depressive disorder: a genome-wide association study. Preprint at MedRxiv. <https://doi.org/10.1101/2020.12.11.20245035>.
 119. GENDEP Investigators; MARS Investigators; and STAR* D Investigators (2013). Common genetic variation and antidepressant efficacy in major depressive disorder: a meta-analysis of three genome-wide pharmacogenetic studies. *Am. J. Psychiat.* 170, 207–217.
 120. Koido, M., Kawakami, E., Fukumura, J., Noguchi, Y., Ohori, M., Nio, Y., Nicoletti, P., Aithal, G.P., Daly, A.K., Watkins, P.B., et al. (2020). Polygenic architecture informs potential vulnerability to drug-induced liver injury. *Nat. Med.* 26, 1541–1548.
 121. Björnsson, E.S. (2015). Drug-induced liver injury: an overview over the most critical compounds. *Arch. Toxicol.* 89, 327–334.
 122. Kaliyaperumal, K., Grove, J.I., Delahay, R.M., Griffiths, W.J., Duckworth, A., and Aithal, G.P. (2018). Pharmacogenomics of drug-induced liver injury (DILI): molecular biology to clinical applications. *J. Hepatol.* 69, 948–957.
 123. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature* 590, 290–299.
 124. Van Hout, C.V., Tachmazidou, I., Backman, J.D., Hoffman, J.D., Liu, D., Pandey, A.K., Gonzaga-Jauregui, C., Khalid, S., Ye, B., Banerjee, N., et al. (2020). Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* 586, 749–756.
 125. Weedon, M., Jackson, L., Harrison, J., Ruth, K., Tyrrell, J., Hattersley, A., and Wright, C. (2021). Use of SNP chips to detect rare pathogenic variants: retrospective, population based diagnostic evaluation. *BMJ* 372, n214.
 126. Ingelman-Sundberg, M., Mkrtchian, S., Zhou, Y., and Lauschke, V.M. (2018). Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genomics* 12, 1–12.
 127. Gordon, A., Fulton, R.S., Qin, X., Mardis, E.R., Nickerson, D.A., and Scherer, S. (2016). PGRNseq: a targeted capture sequencing panel for pharmacogenetic research and implementation. *Pharmacogenetics and Genomics* 26, 161–168.
 128. Rasmussen-Torvik, L.J., Stallings, S.C., Gordon, A.S., Almoquera, B., Basford, M.A., Bielinski, S.J., Brautbar, A., Brilliant, M., Carrell, D.S., Connolly, J., et al. (2014). Design and anticipated outcomes of the eMERGE-PGx project: a multicenter pilot for preemptive pharmacogenomics in electronic health record systems. *Clin. Pharmacol. Ther.* 96, 482–489.
 129. Bush, W.S., Crosslin, D.R., Owusu-Obeng, A., Wallace, J., Almoquera, B., Basford, M.A., Bielinski, S.J., Carrell, D.S., Connolly, J.J., Crawford, D., et al. (2016). Genetic variation among 82 pharmacogenes: the PGRNseq data from the eMERGE network. *Clin. Pharmacol. Ther.* 100, 160–169.
 130. Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., Jean, P.S., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.-A., Fraser, D., et al. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 337, 100–104.
 131. Fujikura, K., Ingelman-Sundberg, M., and Lauschke, V.M. (2015). Genetic variation in the human cytochrome P450 supergene family. *Pharmacogenetics Genom.* 25, 584–594.
 132. Kozyra, M., Ingelman-Sundberg, M., and Lauschke, V.M. (2017). Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet. Med.* 19, 20–29.
 133. Wright, G., Carleton, B., Hayden, M., and Ross, C. (2018). The global spectrum of protein-coding pharmacogenomic diversity. *Pharmacogenomics J.* 18, 187–195.
 134. Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* 95, 5–23.

135. Niinuma, Y., Saito, T., Takahashi, M., Tsukada, C., Ito, M., Hirasawa, N., and Hiratsuka, M. (2014). Functional characterization of 32 CYP2C9 allelic variants. *Pharmacogenomics J.* *14*, 107–114.
136. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* *89*, 82–93.
137. Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., Team, E.L.P., Christiani, D.C., Wurfel, M.M., Lin, X., et al. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* *91*, 224–237.
138. Gloss, B.S., and Dinger, M.E. (2018). Realizing the significance of noncoding functionality in clinical genomics. *Exp. Mol. Med.* *50*, 1–8.
139. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H., and Goldstein, D.B. (2010). Rare variants create synthetic genome-wide associations. *PLoS Biol.* *8*, e1000294.
140. Park, J., Lee, S.Y., Baik, S.Y., Park, C.H., Yoon, J.H., Ryu, B.Y., and Kim, J.H. (2020). Gene-wise burden of coding variants correlates to noncoding pharmacogenetic risk variants. *Int. J. Mol. Sci.* *21*, 3091.
141. Wray, N.R., Purcell, S.M., and Visscher, P.M. (2011). Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol.* *9*, e1000579.
142. Hu, W., Jiang, C., Guan, D., Dierickx, P., Zhang, R., Moscati, A., Nadkarni, G.N., Steger, D.J., Loos, R.J., Hu, C., et al. (2019). Patient adipose stem cell-derived adipocytes reveal genetic variation that predicts antidiabetic drug response. *Cell Stem Cell* *24*, 299–308.e6.
143. Quintanilla Rodriguez, B.S., and Correa, R. (2021). Rosiglitazone. In *StatPearls* (StatPearls Publishing Treasure Island (FL)).
144. Lehmann, J.M., Moore, L.B., Smith-Oliver, T.A., Wilkison, W.O., Willson, T.M., and Kliewer, S.A. (1995). An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor γ (PPAR γ). *J. Biol. Chem.* *270*, 12953–12956.
145. Akiyama, T.E., Sakai, S., Lambert, G., Nicol, C.J., Matsusue, K., Pimprale, S., Lee, Y.-H., Ricote, M., Glass, C.K., Brewer, H.B., et al. (2002). Conditional disruption of the peroxisome proliferator-activated receptor γ gene in mice results in lowered expression of ABCA1, ABCG1, and apoE in macrophages and reduced cholesterol efflux. *Mol. Cell Biol.* *22*, 2607–2619.
146. Wang, Y., Fang, Z., Hong, M., Yang, D., and Xie, W. (2020). Long-noncoding RNAs (lncRNAs) in drug metabolism and disposition, implications in cancer chemo-resistance. *Acta Pharmaceutica Sinica B* *10*, 105–112.
147. Chen, L., Bao, Y., Piekos, S.C., Zhu, K., Zhang, L., and Zhong, X.-b. (2018). A transcriptional regulatory network containing nuclear receptors and long noncoding RNAs controls basal and drug-induced expression of cytochrome P450s in HepaRG cells. *Mol. Pharmacol.* *94*, 749–759.
148. Wang, Y., Yan, L., Liu, J., Chen, S., Liu, G., Nie, Y., Wang, P., Yang, W., Chen, L., Zhong, X., et al. (2019). The HNF1 α -regulated lncRNA HNF1 α -AS1 is involved in the regulation of cytochrome P450 expression in human liver tissues and Huh7 cells. *J. Pharmacol. Exp. Ther.* *368*, 353–362.
149. Yang, Y., Botton, M.R., Scott, E.R., and Scott, S.A. (2017). Sequencing the CYP2D6 gene: from variant allele discovery to clinical pharmacogenetic testing. *Pharmacogenomics* *18*, 673–685.
150. Ammar, R., Paton, T.A., Torti, D., Shlien, A., and Bader, G.D. (2015). Long read nanopore sequencing for detection of HLA and CYP2D6 variants and haplotypes. *F1000Res.* *4*, 17.
151. Qiao, W., Yang, Y., Sebra, R., Mendiratta, G., Gaedigk, A., Desnick, R.J., and Scott, S.A. (2016). Long-read single molecule real-time full gene sequencing of cytochrome P450-2D6. *Hum. Mutat.* *37*, 315–323.
152. Buermans, H.P., Vossen, R.H., Anvar, S.Y., Allard, W.G., Guchelaar, H.-J., White, S.J., den Dunnen, J.T., Swen, J.J., and van der Straaten, T. (2017). Flexible and scalable full-length CYP2D6 long amplicon PacBio sequencing. *Hum. Mutat.* *38*, 310–316.
153. Liau, Y., Maggo, S., Miller, A.L., Pearson, J.F., Kennedy, M.A., and Cree, S.L. (2019). Nanopore sequencing of the pharmacogene CYP2D6 allows simultaneous haplotyping and detection of duplications. *Pharmacogenomics* *20*, 1033–1047.
154. Fukunaga, K., Hishinuma, E., Hiratsuka, M., Kato, K., Okusaka, T., Saito, T., Ikeda, M., Yoshida, T., Zembutsu, H., Iwata, N., et al. (2021). Determination of novel CYP2D6 haplotype using the targeted sequencing followed by the long-read sequencing and the functional characterization in the Japanese population. *J. Hum. Genet.* *66*, 139–149.
155. De Coster, W., Weissensteiner, M.H., and Sedlazeck, F.J. (2021). Towards population-scale long-read sequencing. *Nat. Rev. Genet.* *22*, 572–587.
156. Nadeau, J.H., and Auwerx, J. (2019). The virtuous cycle of human genetics and mouse models in drug discovery. *Nat. Rev. Drug Discov.* *18*, 255–272.
157. Fattinger, K., Roos, M., Vergères, P., Holenstein, C., Kind, B., Masche, U., Stocker, D.N., Braunschweig, S., Kullak-Ublick, G.A., Galeazzi, R.L., et al. (2000). Epidemiology of drug exposure and adverse drug reactions in two Swiss departments of internal medicine. *Br. J. Clin. Pharmacol.* *49*, 158–167.
158. Franconi, F., and Campesi, I. (2014). Pharmacogenomics, pharmacokinetics and pharmacodynamics: interaction with biological differences between men and women. *Br. J. Pharmacol.* *171*, 580–594.
159. Khrantsova, E.A., Davis, L.K., and Stranger, B.E. (2019). The role of sex in the genomics of human complex traits. *Nat. Rev. Genet.* *20*, 173–190.
160. Marengoni, A., Pasina, L., Concoreggi, C., Martini, G., Brognoli, F., Nobili, A., Onder, G., and Bettoni, D. (2014). Understanding adverse drug reactions in older adults through drug-drug interactions. *Eur. J. Intern. Med.* *25*, 843–846.
161. Malki, M.A., and Pearson, E.R. (2020). Drug-drug-gene interactions and adverse drug reactions. *Pharmacogenomics J.* *20*, 355–366.
162. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* *337*, 816–821.
163. Karlgren, M., Simoff, I., Keiser, M., Oswald, S., and Artursson, P. (2018). CRISPR-Cas9: a new addition to the drug metabolism and disposition tool box. *Drug Metab. Dispos.* *46*, 1776–1786.
164. Wang, D., Poi, M.J., Sun, X., Gaedigk, A., Leeder, J.S., and Sadee, W. (2014). Common CYP2D6 polymorphisms affecting alternative splicing and transcription: long-range haplotypes with two regulatory variants modulate CYP2D6 activity. *Hum. Mol. Genet.* *23*, 268–278.

165. Wang, D., Papp, A.C., and Sun, X. (2015). Functional characterization of CYP2D6 enhancer polymorphisms. *Hum. Mol. Genet.* *24*, 1556–1562.
166. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* *31*, 822–826.
167. Nelson, D.R., and Nebert, D.W. (2011). Cytochrome P450 (CYP) gene superfamily. *Encyclopedia Life Sci* *0*, 1–13.
168. Cheetham, S.W., Faulkner, G.J., and Dinger, M.E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nat. Rev. Genet.* *21*, 191–201.
169. Kantor, E.D., Rehm, C.D., Haas, J.S., Chan, A.T., and Giovannucci, E.L. (2015). Trends in prescription drug use among adults in the United States from 1999–2012. *JAMA* *314*, 1818–1831.
170. McDermott, U. (2019). Large-scale compound screens and pharmacogenomic interactions in cancer. *Curr. Opin. Genet. Dev.* *54*, 12–16.
171. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. *Nat. Methods* *11*, 801–807.
172. Amorosi, C.J., Chiasson, M.A., McDonald, M.G., Wong, L.H., Sitko, K.A., Boyle, G., Kowalski, J.P., Rettie, A.E., Fowler, D.M., and Dunham, M.J. (2021). Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. *Am. J. Hum. Genet.* *108*, 1735–1751.
173. Swanzey, E., O'Connor, C., and Reinholdt, L.G. (2021). Mouse genetic reference populations: cellular platforms for integrative systems genetics. *Trends Genet.* *37*, 251–265.
174. Broman, K.W. (2005). The genomes of recombinant inbred lines. *Genetics* *169*, 1133–1146.
175. Kim, J., Koo, B.-K., and Knoblich, J.A. (2020). Human organoids: model systems for human biology and medicine. *Nat. Rev. Mol. Cell Biol.* *21*, 571–584.
176. Dekkers, J.F., Berkers, G., Kruisselbrink, E., Vonk, A., De Jonge, H.R., Janssens, H.M., Bronsveld, I., van de Graaf, E.A., Nieuwenhuis, E.E., Houwen, R.H., et al. (2016). Characterizing responses to CFTR-modulating drugs using rectal organoids derived from subjects with cystic fibrosis. *Sci. Transl. Med.* *8*, 344ra84.
177. Berkers, G., van Mourik, P., Vonk, A.M., Kruisselbrink, E., Dekkers, J.F., de Winter-de Groot, K.M., Arets, H.G., Marckvan der Wilt, R.E., Dijkema, J.S., Vanderschuren, M.M., et al. (2019). Rectal organoids enable personalized treatment of cystic fibrosis. *Cell Rep.* *26*, 1701–1708.e3.
178. Pound, P., and Bracken, M.B. (2014). Is animal research sufficiently evidence based to be a cornerstone of biomedical research? *BMJ* *348*, g3387.
179. Van Norman, G.A. (2019). Limitations of animal studies for predicting toxicity in clinical trials: is it time to rethink our current approach? *JACC: Basic Translational Sci.* *4*, 845–854.
180. Nair, S., Hsu, D., and Celi, L.A. (2016). Challenges and opportunities in secondary analyses of electronic health record data. In *Secondary Analysis of Electronic Health Records* (Springer), pp. 17–26.
181. Dinh-Le, C., Chuang, R., Chokshi, S., and Mann, D. (2019). Wearable health technology and electronic health record integration: scoping review and future directions. *JMIR mHealth and uHealth* *7*, e12861.
182. Robinson, J.R., Denny, J.C., Roden, D.M., and Van Driest, S.L. (2018). Genome-wide and phenome-wide approaches to understand variable drug actions in electronic health records. *Clin. Transl. Sci.* *11*, 112–122.
183. Tasa, T., Krebs, K., Kals, M., Mägi, R., Lauschke, V.M., Haller, T., Puurand, T., Remm, M., Esko, T., Metspalu, A., et al. (2019). Genetic variation in the Estonian population: pharmacogenomics study of adverse drug effects using electronic health records. *Eur. J. Hum. Genet.* *27*, 442–454.
184. McInnes, G., and Altman, R.B. (2021). Drug response pharmacogenetics for 200,000 UK Biobank participants. *Pac. Symp. Biocomput.* *26*, 184–195.
185. Xu, D., Wang, C., Khan, A., Shang, N., He, Z., Gordon, A., Kullo, I.J., Murphy, S., Ni, Y., Wei, W.-Q., et al. (2021). Quantitative disease risk scores from EHR with applications to clinical risk stratification and genetic studies. *NPJ Digital Med.* *4*, 116.
186. Li, R., Chen, Y., Ritchie, M.D., and Moore, J.H. (2020). Electronic health records and polygenic risk scores for predicting disease risk. *Nat. Rev. Genet.* *21*, 493–502.
187. Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., et al. (2015). Cohort profile: Estonian biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* *44*, 1137–1147.
188. Leitsalu, L., Alavere, H., Tammesoo, M.-L., Leego, E., and Metspalu, A. (2015). Linking a population biobank with national health registries—the Estonian experience. *J. Personalized Med.* *5*, 96–106.
189. Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W.A., Li, R., Manolio, T.A., Sanderson, S.C., Kannry, J., Zinberg, R., Basford, M.A., et al. (2013). The electronic medical records and genomics (eMERGE) network: past, present, and future. *Genet. Med.* *15*, 761–771.
190. Elliott, P., and Peakman, T.C. (2008). The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int. J. Epidemiol.* *37*, 234–244.
191. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* *562*, 203–209.
192. Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C., Liu, D., Locke, A.E., Balasubramanian, S., et al. (2021). Exome sequencing and analysis of 454,787 UK biobank participants. *Nature* *599*, 628–634.
193. Carey, D.J., Fetterolf, S.N., Davis, F.D., Faucett, W.A., Kirchner, H.L., Mirshahi, U., Murray, M.F., Smelser, D.T., Gerhard, G.S., and Ledbetter, D.H. (2016). The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet. Med.* *18*, 906–913.
194. Dewey, F.E., Murray, M.F., Overton, J.D., Habegger, L., Leader, J.B., Fetterolf, S.N., O'Dushlaine, C., Van Hout, C.V., Staples, J., Gonzaga-Jauregui, C., et al. (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* *354*, aaf6814.
195. Gaziano, J.M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., et al. (2016). Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* *70*, 214–223.
196. Hunter-Zinck, H., Shi, Y., Li, M., Gorman, B.R., Ji, S.-G., Sun, N., Webster, T., Liem, A., Hsieh, P., Devineni, P., et al. (2020). Genotyping array design and data quality control in the Million Veteran Program. *Am. J. Hum. Genet.* *106*, 535–548.

197. Wei, C.-Y., Yang, J.-H., Yeh, E.-C., Tsai, M.-F., Kao, H.-J., Lo, C.-Z., Chang, L.-P., Lin, W.-J., Hsieh, F.-J., Belsare, S., et al. (2021). Genetic profiles of 103,106 individuals in the Taiwan Biobank provide insights into the health and history of Han Chinese. *NPJ Genomic Med.* *6*, 1–10.
198. Lin, J.-C., Fan, C.-T., Liao, C.-C., and Chen, Y.-S. (2018). Taiwan Biobank: making cross-database convergence possible in the Big Data era. *Gigascience* *7*, 1–4.
199. H3Africa Consortium (2014). Enabling the genomic revolution in Africa. *Science* *344*, 1346–1348.
200. Mulder, N., Abimiku, A., Adebamowo, S.N., de Vries, J., Matimba, A., Olowoyo, P., Ramsay, M., Skelton, M., and Stein, D.J. (2018). H3Africa: current perspectives. *Pharmacogenomics Personalized Med.* *11*, 59–66.
201. Tadaka, S., Hishinuma, E., Komaki, S., Motoike, I.N., Kawashima, J., Saigusa, D., Inoue, J., Takayama, J., Okamura, Y., Aoki, Y., et al. (2021). jMorp updates in 2020: large enhancement of multi-omics data resources on the general Japanese population. *Nucleic Acids Res.* *49*, D536–D544.
202. Ido, K., Nakamura, N., and Nakayama, M. (2019). Miyagi medical and welfare information network: a backup system for patient clinical information after the great east Japan earthquake and tsunami. *Tohoku J. Exp. Med.* *248*, 19–25.
203. Locke, A.E., Steinberg, K.M., Chiang, C.W., Service, S.K., Havulinna, A.S., Stell, L., Pirinen, M., Abel, H.J., Chiang, C.C., Fulton, R.S., et al. (2019). Exome sequencing of Finnish isolates enhances rare-variant association power. *Nature* *575*, E4.
204. The All of Us Research Program Investigators (2019). The “All of Us” research program. *New Engl. J. Med.* *381*, 668–676.
205. UK Biobank. (2020). UK Biobank Website. <https://www.ukbiobank.ac.uk/>.
206. van der Wouden, C.H., Cambon-Thomsen, A., Cecchin, E., Cheung, K.-C., Dávila-Fajardo, C.L., Deneer, V.H., et al. (2017). Implementing pharmacogenomics in Europe: design and implementation strategy of the Ubiquitous Pharmacogenomics Consortium. *Clin. Pharmacol. Ther.* *102*, 152.
207. Chumnumwat, S., Lu, Z.H., Sukasem, C., Winther, M.D., Capule, F.R., Hamid, A.A.A., et al. (2019). Southeast Asian Pharmacogenomics research network (SEAPharm): current status and perspectives. *Public Health Genomics* *22*, 132–139.
208. Runcharoen, C., Fukunaga, K., Sensorn, I., Iemwimangsa, N., Klumsathian, S., Tong, H., et al. (2021). Prevalence of pharmacogenomic variants in 100 pharmacogenes among southeast asian populations under the collaboration of the southeast asian pharmacogenomics research network (SEAPharm). *Hum. Genome Var.* *8*, 7.
209. Mpye, K., Matimba, A., Dzobo, K., Chirikure, S., Wonkam, A., and Dandara, C. (2017). Disease burden and the role of pharmacogenomics in African populations. *Glob. Health Epidemiol. Genomics* *2*, e1.
210. Wendt, F.R., Koller, D., Pathak, G.A., Jacoby, D., Miller, E.J., and Polimanti, R. (2021). Biobank scale pharmacogenomics informs the genetic underpinnings of simvastatin use. *Clin. Pharmacol. Ther.* *110*, 777–785.
211. Manolio, T.A., Goodhand, P., and Ginsburg, G. (2020). The International Hundred Thousand Plus Cohort Consortium: integrating large-scale cohorts to address global scientific challenges. *The Lancet Digital Health* *2*, e567–e568.
212. Anglemyer, A., Horvath, H.T., and Bero, L. (2014). Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst. Rev.* *2014*, MR000034.
213. Davey Smith, G., and Ebrahim, S. (2003). ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* *32*, 1–22.
214. Bowden, J., Pilling, L.C., Türkmen, D., Kuo, C.-L., and Melzer, D. (2021). The Triangulation Within a Study (TWIST) framework for causal inference within pharmacogenetic research. *PLoS Genet.* *17*, e1009783.
215. Ramsey, L.B., Johnson, S.G., Caudle, K.E., Haidar, C.E., Voora, D., Wilke, R.A., Maxwell, W.D., McLeod, H.L., Krauss, R.M., Roden, D.M., et al. (2014). The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. *Clin. Pharmacol. Ther.* *96*, 423–428.
216. Türkmen, D., Masoli, J.A., Kuo, C.-L., Bowden, J., Melzer, D., and Pilling, L.C. (2022). Statin treatment effectiveness and the SLCO1B1*5 reduced function genotype: long-term outcomes in women and men. *Br. J. Clin. Pharmacol.* <https://doi.org/10.1111/bcp.15245>.
217. Sheikhalishahi, S., Miotto, R., Dudley, J.T., Lavelli, A., Rinaldi, F., and Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med. Inform.* *7*, e12239.
218. Shi, L., and Wang, Z. (2019). Computational strategies for scalable genomics analysis. *Genes* *10*, 1017.
219. Chevrier, R., Foufi, V., Gaudet-Blavignac, C., Robert, A., and Lovis, C. (2019). Use and understanding of anonymization and de-identification in the biomedical literature: scoping review. *J. Med. Internet Res.* *21*, e13484.
220. Manzoni, C., Kia, D.A., Vandrovцова, J., Hardy, J., Wood, N.W., Lewis, P.A., and Ferrari, R. (2018). Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Brief. Bioinform.* *19*, 286–302.
221. Porcu, E., Sjaarda, J., Lepik, K., Carmeli, C., Darrous, L., Sulc, J., Mounier, N., and Kutalik, Z. (2021). Causal inference methods to integrate omics and complex traits. *Cold Spring Harbor Perspect. Med.* *11*, a040493.
222. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* *10*, e1004383.
223. Hormozdiari, F., Van De Bunt, M., Segre, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankaraman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* *99*, 1245–1260.
224. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* *47*, 1091–1098.
225. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., De Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* *48*, 245–252.
226. Porcu, E., Rüeger, S., Lepik, K., Santoni, F.A., Reymond, A., and Kutalik, Z. (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat. Commun.* *10*, 3300.
227. Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P.,

- et al. (2018). Genomic atlas of the human plasma proteome. *Nature* 558, 73–79.
228. Zhao, H., Rasheed, H., Nøst, T.H., Cho, Y., Liu, Y., Bhatta, L., et al. (2022). Proteome-wide mendelian randomization in global biobank meta-analysis reveals multi-ancestry drug targets for common diseases. Preprint at MedRxiv. <https://doi.org/10.1101/2022.01.09.21268473>.
 229. Liu, D., Wang, Y., Jing, H., Meng, Q., and Yang, J. (2021). Mendelian randomization integrating GWAS and mQTL data identified novel pleiotropic DNA methylation loci for neuropathology of Alzheimer's disease. *Neurobiol. Aging* 97, 18–27.
 230. Moayyeri, A., Cheung, C.-L., Tan, K.C., Morris, J.A., Cerani, A., Mohney, R.P., Richards, J.B., Hammond, C., Spector, T.D., and Menni, C. (2018). Metabolomic pathways to osteoporosis in middle-aged women: a genome-metabolome-wide Mendelian randomization study. *J. Bone Mineral Res.* 33, 643–650.
 231. Denny, J.C., Van Driest, S.L., Wei, W.-Q., and Roden, D.M. (2018). The influence of big (clinical) data and genomics on precision medicine and drug development. *Clin. Pharmacol. Ther.* 103, 409–418.
 232. Lin, E., and Lane, H.-Y. (2017). Machine learning and systems genomics approaches for multi-omics data. *Biomarker Res.* 5, 2.
 233. Athreya, A.P., Neavin, D., Carrillo-Roa, T., Skime, M., Bieracka, J., Frye, M.A., Rush, A.J., Wang, L., Binder, E.B., Iyer, R.K., et al. (2019). Pharmacogenomics-driven prediction of antidepressant treatment outcomes: a machine-learning approach with multi-trial replication. *Clin. Pharmacol. Ther.* 106, 855–865.
 234. Lin, E., Lin, C.-H., and Lane, H.-Y. (2020). Precision psychiatry applications with pharmacogenomics: artificial intelligence and machine learning approaches. *Int. J. Mol. Sci.* 21, 969.
 235. Conesa, A., and Beck, S. (2019). Making multi-omics data accessible to researchers. *Sci. Data* 6, 251.
 236. Burton, J., Bhattacharya, S., Romero, K., and Conrado, D.J. (2020). Open data for clinical pharmacology. *Clin. Pharmacol. Ther.* 107, 703–706.
 237. Musa, A., Ghoraie, L.S., Zhang, S.-D., Glazko, G., Yli-Harja, O., Dehmer, M., Haibe-Kains, B., and Emmert-Streib, F. (2018). A review of Connectivity Map and computational approaches in pharmacogenomics. *Brief. Bioinform.* 19, 506–523.
 238. Misra, B.B., Langefeld, C., Olivier, M., and Cox, L.A. (2019). Integrated omics: tools, advances and future approaches. *J. Mol. Endocrinol.* 62, R21–R45.
 239. Ruiz, C., Zitnik, M., and Leskovec, J. (2021). Identification of disease treatment mechanisms through the multiscale interactome. *Nat. Commun.* 12, 1796.
 240. Lauschke, V.M., and Ingelman-Sundberg, M. (2020). Emerging strategies to bridge the gap between pharmacogenomic research and its clinical implementation. *NPJ Genomic Med.* 5, 9.
 241. Zhou, Y., Mkrtychian, S., Kumondai, M., Hiratsuka, M., and Lauschke, V.M. (2019). An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J.* 19, 115–126.
 242. McInnes, G., Dalton, R., Sangkuhl, K., Whirl-Carrillo, M., Lee, S.-b., Tsao, P.S., Gaedigk, A., Altman, R.B., and Woodahl, E.L. (2020). Transfer learning enables prediction of CYP2D6 haplotype function. *PLoS Comput. Biol.* 16, e1008399.
 243. van der Lee, M., Allard, W.G., Vossen, R.H.A.M., Baak-Pablo, R.F., Menafra, R., Deiman, B.A.L.M., Deenen, M.J., Neven, P., Johansson, I., Gastaldello, S., et al. (2021). Toward predicting CYP2D6-mediated variable drug response from CYP2D6 gene sequencing data. *Sci. Transl. Med.* 13, eabf3637.
 244. So, H.-C., Chau, C.K.-L., Chiu, W.-T., Ho, K.-S., Lo, C.-P., Yim, S.H.-Y., and Sham, P.-C. (2017). Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat. Neurosci.* 20, 1342–1349.
 245. Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell* 171, 1437–1452.e17.
 246. Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Félix, E., Magariños, M.P., Mosquera, J.F., Mutowo, P., Nowotka, M., et al. (2019). ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* 47, D930–D940.
 247. Pushpakom, S., Iorio, F., Evers, P.A., Escott, K.J., Hopper, S., Wells, A., Doig, A., Guillelliams, T., Latimer, J., McNamee, C., et al. (2019). Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18, 41–58.
 248. Stéphanou, A., Fanchon, E., Innominato, P.F., and Ballesta, A. (2018). Systems biology, systems medicine, systems pharmacology: the what and the why. *Acta Biotheor.* 66, 345–365.
 249. Nielsen, J. (2017). Systems biology of metabolism. *Annu. Rev. Biochem.* 86, 245–275.
 250. Nagar, S., Argikar, U.A., and Tweedie, D.J. (2014). Enzyme kinetics in drug metabolism: fundamentals and applications. In *Enzyme Kinetics in Drug Metabolism* (Springer), pp. 1–6.
 251. Salvy, P., and Hatzimanikatis, V. (2020). The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat. Commun.* 11, 30.
 252. Brunk, E., Sahoo, S., Zielinski, D.C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Gonzalez, G.A.P., Aurich, M.K., et al. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* 36, 272–281.
 253. Zhang, J.D., Sach-Peltason, L., Kramer, C., Wang, K., and Ebeling, M. (2020). Multiscale modelling of drug mechanism and safety. *Drug Discov. Today* 25, 519–534.
 254. Chenoweth, M.J., Giacomini, K.M., Pirmohamed, M., Hill, S.L., van Schaik, R.H., Schwab, M., Shuldiner, A.R., Relling, M.V., and Tyndale, R.F. (2020). Global pharmacogenomics within precision medicine: challenges and opportunities. *Clin. Pharmacol. Ther.* 107, 57–61.
 255. Nelson, M.R., Johnson, T., Warren, L., Hughes, A.R., Chisoe, S.L., Xu, C.-F., and Waterworth, D.M. (2016). The genetics of drug efficacy: opportunities and challenges. *Nat. Rev. Genet.* 17, 197–206.
 256. Bentley, A.R., Callier, S., and Rotimi, C.N. (2017). Diversity and inclusion in genomic research: why the uneven progress? *J. Community Genet.* 8, 255–266.
 257. Knepper, T.C., and McLeod, H.L. (2018). When will clinical trials finally reflect diversity? *Nature* 557, 157–159.
 258. Mills, M.C., and Rahal, C. (2019). A scientometric review of genome-wide association studies. *Commun. Biol.* 2, 9.
 259. Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The missing diversity in human genetic studies. *Cell* 177, 26–31.
 260. Gaedigk, A., Sangkuhl, K., Whirl-Carrillo, M., Klein, T., and Leeder, J.S. (2017). Prediction of CYP2D6 phenotype from genotype across world populations. *Genet. Med.* 19, 69–76.