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 Kutalik2,3,4,*

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 effective 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
reactions are under genetic control. 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). 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.” The intuition was confirmed decades later, as studies revealed a higher prevalence of causative glucose-6-phosphate dehydrogenase deficiency in Africans, 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. 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 debrisoquine and the sodium channel blocker sparteine 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. As oxidative metabolism of other drugs seemed unaffected, 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. For instance, the etiology of the debrisoquine and sparteine non-metabolizer phenotype was elucidated by purifying the relevant CYP and cloning the relevant alleles. 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, 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. Suspcion 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.

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, 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. 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. Hence, many associations failed to replicate and further proof was warranted prior to clinical implementation.
Randomized controlled trials

Randomized controlled trials (RCTs) have become the gold standard of evidence-based medicine over the course of the 20th century. 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. 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, 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. 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...
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. Later, microarray studies revealed population differentiation across ADME genes and sequencing studies generated global maps of PGx allele frequencies. Following, a gene-based population differentiation method that optimally accounts for rare variants showed that 10% of investigated PGx genes exhibit high population differentiation.

What are the consequences of lack of diversity?

Despite growing awareness around the importance of inclusiveness, the majority of genetic studies remain Euro-centric, 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. 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. 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. 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. 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.

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). 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). 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). 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.

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. Later, microarray studies revealed population differentiation across ADME genes and sequencing studies generated global maps of PGx allele frequencies. Following, a gene-based population differentiation method that optimally accounts for rare variants showed that 10% of investigated PGx genes exhibit high population differentiation.

What are the consequences of lack of diversity?

Despite growing awareness around the importance of inclusiveness, the majority of genetic studies remain Euro-centric, 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. 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. 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. 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. 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.

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). 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). 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). 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), prompting the establishment of clinical screening guidelines. 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.

RCTs provide strong support of causality between an exposure and an outcome, but their role in PGx has been called into question. 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, 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). Clopidogrel is the most widely prescribed anti-platelet medication, but due to genetically driven variability in drug response, it is suspected to increase the risk for MACEs. As CYP2C19 (MIM: 124020, 609535) is required to activate clopidogrel, CYP2C19 loss-of-function (LOF) carriers in the genotype-guided group were prescribed ticagrelor instead of clopidogrel. 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. Another limitation is that only two LOF alleles were considered, although recent evidence demonstrated a polygenic architecture of clopidogrel response (Polygenic pharmacogenomics). 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
that CYP2C19-guided prescription could identify patients benefiting from alternative anti-platelet medication,\textsuperscript{53,54} prompting the CPIC to update its recommendations in 2022, thereby providing guidance to clinicians on how to interpret CYP2C19 genetic test results.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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-genetic tests, and most physicians unconsciously practice PGx by adjusting patient’s treatments.\textsuperscript{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\textsuperscript{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.\textsuperscript{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 pharmaco genomics. Conceptually similar and therefore both referred to as PGx hereafter, pharmaco genomics 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
known to encode a hepatic statin transporter, 87 it had cogenetics captured, and inter-connected resources. Search terms more it will rely on the presence of centralized, well-struc-
tional 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 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 × 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
detect PGx interactions. One of the first drug responses investigated by GWAS was warfarin maintenance dose. At the time, warfarin was the most widely used oral anticoagulant, with prescription rates totaling 0.5%–1.5% of the population. 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. 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. 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. A second GWAS with improved power (n = 1,053) identified significant signals for VKORC1, CYP2C9, and CYP4F2 (MIM: 604426). 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. 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). Arrival of direct-acting antivirals improved treatment success rate, especially in patients with unfavorable IFNL3 genotype. Nowadays, new generations of well-tolerated and highly effective direct-acting antiviral regimens bypass the need for pegylated interferons altogether. 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, 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 GWAS.

### 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. Since, the polygenic nature of complex traits has become widely accepted and pushed to the extreme with the omicnic model, which hypothesizes that all genes expressed within a phenotype-relevant cell type contribute to the phenotype in question. 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, but these loci were not sufficient to explain the estimated 34% heritability of the drug response. Instead, chromosome-wise heritability estimates pointed at the presence of multiple small-effect variants spread across the genome. In the following years, independent, better-powered GWASs implicated an increasing number of genes in metformin response.
One of the yield mixed results, as highlighted in a 2021 comprehensive review of 51 PGx PGS studies. So far, applications in PGx are sparse and have yet to fully reveal 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. 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) that improve imputation quality across diverse ancestries. Lastly, microarrays are not suited to accurately assess rare variants, despite evidence suggesting that these might account for some of the missing heritability in PGx traits.

**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, 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, will assess PGx variation in the electronic Medical Records and Genomics (eMERGE)-PGx project (n = 9,000). Another study (n = 14,002) analyzed rare variants in 202 drug target genes, suggesting that rare variants are both numerous and population specific. Genetic characterization of the 57 human CYP genes revealed over 6,000 new SNVs. 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. \textsuperscript{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). \textsuperscript{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). \textsuperscript{137}

---

to the conclusion that >90% of SNVs in these genes are rare. \textsuperscript{132,133} Despite their low frequency, rare variants strongly contribute to functional variability, \textsuperscript{126,132} many being predicted to be deleterious. \textsuperscript{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 testing \textsuperscript{134} (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). \textsuperscript{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. \textsuperscript{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. \textsuperscript{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. \textsuperscript{139} Drug-response-associated non-coding variants could simply tag the deleterious mutational burden in corresponding PGx genes, \textsuperscript{140} even if it has been argued that synthetic associations created by rare variants are unlikely to explain most GWAS results. \textsuperscript{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, \textsuperscript{142} an antidiabetic drug associated with adverse effects, such as increased cholesterol levels. \textsuperscript{143} Rosiglitazone mediates its therapeutic effect by binding and activating the nuclear receptor PPAR\(_\gamma\) (MIM: 601487), \textsuperscript{144} which induces the expression of multiple genes, including the cholesterol efflux transporter *ABCA1* (MIM: 600046). \textsuperscript{145} Chromatin immunoprecipitation sequencing showed that A/A carriers lack a rosiglitazone-induced PPAR\(_\gamma\)-binding site and fail to induce expression of nearby *ABCA1* upon administration of the drug. \textsuperscript{142} Confirming the clinical relevance of the variant, diabetic patients with the corresponding genotype experienced a reduced rosiglitazone-induced increase in cholesterol levels. \textsuperscript{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. \textsuperscript{146} While this area of research remains in its infancy, it was shown that the transcription factor HNF1\(_z\) (MIM: 142410) and the lncRNA HNF1\(_z\)-AS1 form a regulatory network that controls the expression of several *CYP* genes. \textsuperscript{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. \textsuperscript{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. \textsuperscript{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, \textsuperscript{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, \textsuperscript{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, \textsuperscript{156} making it possible to establish causal
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) 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. 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. CRISPR technologies have been linked to off-target mutagenesis at sites with high sequence similarity—which are particularly common among ADME genes—even if careful usage of the tool should mitigate this issue. More importantly, the large number of PGx variants and persistent increase in drug prescription and polypharmacy 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. 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. 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 experiment and was recently applied to generate abundance and activity scores for 87% of the possible missense variants in CYP2C9. 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). Producing genetically identical individuals, they represent an 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 PGx study. 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. 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 PGx study. 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. While widespread clinical implementation awaits, it was shown that organoids from cystic fibrosis patients can predict drug efficacy.

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, but the virtuous cycle between human observational studies and in vitro and in vivo experimentation in model organisms 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. EHRs commonly encompass patient demographics, medical history, drug prescriptions, and in some cases laboratory results, radiological images, and wearable device data. Integration of EHRs is key to the realization of precision medicine, and current research already benefits from it by conducting nonrandomized studies. 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.

Electronic health records in pharmacogenomics
Increasingly, biobanks match genotype entries with EHRs, broadening the phenotype query space exploitable for PGx research. 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. FinnGen plans to recruit ~500,000 individuals by 2023 with access to EHRs from national health registries comprehending medication usage. Similarly, the Estonian Biobank contains detailed information about drug purchase and disease incidences collected since 2000. In parallel, many initiatives focus on the clinical comprehension of PGx, such as the American eMERGE-PGx project, the European Ubiquitous Pharmacogenomics (U-PGx) consortium, the South East Asian Pharmacogenomics Research Network (SEAPharm), or the African Pharmacogenomics Consortium (APC). 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 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. Another UKB study discovered an association between statin usage and NAT2 (MIM: 612182) genotype. The first study making use of WGS-coupled EHRs was conducted in ~2,500 Estonian Biobank participants. Increasing sample size to ~16,000 by including individuals with imputed genotypes, six known and nine new PGx interactions were identified. 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. 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, 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, 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, and randomization between study participants is mimicked based on known observed covariates. Alternatively, Mendelian randomization (MR) approaches account for unobserved confounders, 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, 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. Applying TWIST to the UKB revealed that CYP2C19 LOF carriers experienced a 0.28% increased risk of stroke (p = 0.048) while on clopidogrel.

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 SLC01B1*5 allele known to impair intracellular statin uptake would have 0.147 mmol/L lower total cholesterol were they treated with a lipid-lowering medication unaffected by their genotype, 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 |
|---|---|---|---|---|---|---|
| **Estonian Genome Project** | 2002–2007 | 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, 188; https://genomics.ut.ee |
| **eMERGE** | 2007–2013 | >135,000 | genome-wide genotyping (n > 105,000); WES (n > 3,700); WGS (n > 1,700); PGBRNseq (n = 9,000) | all participants are linked to EHRs | None | Gottesman et al. 189, Bush et al. 129; https://emerge-network.org |
| **UK Biobank** | 2006–2010 | 500,000 | genome-wide genotyping (n = 500,000); WES (n = 500,000); WGS (n = 500,000) | n = 230,000 | blood biomarkers (n = 500,000); proteomics (n = 15,000); 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 |
| **DiscovEHR** | 2007–2013 | >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.discoverehshare.com |
| **Million Veteran Program** | 2011–2021 | >1,000,000 | genome-wide genotyping (n = 1,000,000); WES; WGS (n = 100,000) | all participants are linked to Veteran Affairs EHR | blood biomarkers; metabolomics; methylomics | Gaziano et al. 195, Hunter-Zinck et al. 196; https://www.mvpg.x.gov |
| **Taiwan Biobank** | 2012–2022 | 200,000 | genome-wide genotyping (n = 200,000); WGS (n = 1,500) | national EHRs available | blood biomarkers; metabolomics | Wei et al. 197, Lin et al. 198; https://www.twbiobank.org.tw |
| **H3Africa** | 2012–2017 | >100,000 | genome-wide genotyping; WES; WGS (n > 400) | None | blood biomarkers; microbiomics | H3Africa Consortium et al. 199, Mulder et al. 200; https://h3africa.org |
| **Tohoku Medical Megabank Project** | 2013–2017 | 150,000 | genome-wide genotyping; 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 |
| **FinnGen** | 2017–2023 | 500,000 | genome-wide genotyping (n = 500,000) | all participants are linked to national EHRs | biological assays | Locke et al. 203; https://www.finngen.fi |
| **All of Us** | 2018–2023 | >1,000,000 | genomics assays | all participants are linked to national EHRs | biological assays | The All of Us Research Program Investigators 204; https://www.researchallofus.org |

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

*Notes:
- eMERGE-PGx Project
- Planned.
- 150,000 samples released in 2021.
- 460,000 samples released in 2020.
- Planned in the population community-based cohort.
- 100,000 samples released in 2021.
- Not yet crosslinked.
- Sample sizes and available measurements are cohort dependent.
- By 2023.
- 350,000 samples released in 2022.
manual review by clinical experts is not scalable, automated solutions based on natural language-processing algorithms are required to extract informative features. Second, the massive size of biobanks necessitates sufficient storage capacities and computational resources to conduct analyses. 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. 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. 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, 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. Similarly, transcriptome-wide association studies (TWASs) pinpoint genes whose differential expression is associated with a given phenotype. 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, proteome, methylome (CpG sites), or metabolome. Not widely employed in PGx, 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 response for antidepressant use, precision psychiatry, and warfarin dosage (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. Furthermore, study results are often siloed and only few databases with unbiased drug-screening experiments are publicly available. There is a need for large-scale in vitro and in vivo perturbation experiments measuring drug effects on multiple omics layers, as results from these assays could open new avenues to study biomarker combinations and drug-drug-gene interactions. 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, 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. System-level PGx is likely to become more elaborate in the future (Perspectives in mechanistic approaches), providing mechanistic explanations for pharmac-o-mics 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. 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. 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 variants 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. 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. Training a convolutional neural network based on star alleles definitions, in vitro enzymatic activity of both known and unknown CYP2D6 haplotypes could be predicted. Similarly, a neural network was trained on a prospective cohort of breast cancer patients to predict CYP2D6-mediated tamoxifen metabolism. 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 GWASs’ 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), a library containing the expression profile of 1,000 genes upon perturbation by ~20,000 small molecules. Although there is no protein analogue to CMap, ChEMBL, the largest database of curated drug targets, and DrugBank form valuable resources to conduct similar studies at the proteome level. Drug repositioning is a topic on its own, 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. Systems biology aims at constructing metabolic network models describing complex biological systems, with nodes representing metabolite concentrations connected by mathematical equations defined by enzyme kinetics. Since enzyme kinetics is at the core of drug metabolism, 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. Efforts to reconstruct a genome-scale network are concerted in Recon3D, the most comprehensive human metabolic network to date. 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
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.

**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. 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.
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.261 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 pharmacogenomics: a case for preemptive genotyping. Clin. Pharmacol. Ther. 92, 235–242.
7. Chanfreau-Coffinier, C., Hull, L.E., Lynch, J.A., DuVall, S.L., Damraper, 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. Pharmacogenomics J. 17, 89–106.
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.
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., Abhijnist, E., Spencer, C.C., Groop, L., Morris, A.D., et al. (2014). Heritability of variation in glycemic 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., Cole- man, R.L., Cavendall, R., Hawley, S.A., Donnelly, L.A., Scho- field, 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., Cavendall, R., Bennett, A.J., Groves, C.J., Coleman, R.L., Van Der Heij- den, 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., Hedderon, M.M., Kubo, M., Her- man, M.A., et al. (2018). Genetic variants in CPA6 and PRPF31 are associated with variation in response to metfor- min in individuals with type 2 diabetes. Diabetes 67, 1428–1440.

113. Zhang, Y., Hu, Y., Ho, K., Hartzell, 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, C., Aka, I.T., Birdwell, K.A., Gordon, A.S., Ro- den, 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., Landor, 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 Drogemöller, B.I. (2021). A systematic review and analysis of the use of polygenic scores in pharma- cogenomics. 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., Marше, V.S., Adams, M.J., et al. (2020). Antidepressant response in major depressive disorder: a genome-wide association study. Pre- print at MedRxiv. https://doi.org/10.1101/2020.12.11. 20245035.

119. GENDEP Investigators; MARS Investigators; and STAR*D Investi- gators (2013). Common genetic variation and anti-de- pressant efficacy in major depressive disorder: a meta-anal- ysis 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. Kalipaperumal, 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., Zpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse ge- nomes 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. Ingelmann-Sundberg, M., Mkrtchian, S., Zhou, Y., and Lauschke, V.M. (2018). Integrating rare genetic variants into pharmacogenetic drug response predictions. Hum. Ge- nomics 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. Pharmacogenomics and Genomics 26, 161–168.

128. Rasmussen-Torvik, L.J., Stallings, S.C., Gordon, A.S., Almo- guera, B., Basford, M.A., Bielinski, S.J., Brautbar, A., Brillant, M., Carrell, D.S., Connolly, J., et al. (2014). Design and antici- pitated 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-O博ng, A., Wallace, J., Al- moguera, B., Basford, M.A., Bielinski, S.J., Carrell, A., 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., Ingelmann-Sundberg, M., and Lauschke, V.M. (2015). Genetic variation in the human cytochrome P450 su- pergene family. Pharmacogenetics Genom. 25, 584–594.

132. Kozyra, M., Ingelmann-Sundberg, M., and Lauschke, V.M. (2017). Rare genetic variants in cellular transporters, meta- bolic enzymes, and nuclear receptors can be important deter- minants 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.
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). Deep mutational scanning and pharmacogenomic interactions in cancer. Curr. Opin. Genet. Dev. 54, 12–16.

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. Swanzye, 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., Kruiselbrink, 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, 344ra64.

177. Berkers, G., van Mourik, P., Vonk, A.M., Kruiselbrink, E., Dekkers, J.F., de Winter-de Groot, K.M., Arets, H.G., Manck-van 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 phenotype-wide approaches to understand variable drug actions in electronic health records. Clin. Transl. Sci. 11, 112–122.

183. Tasa, T., Krebs, K., Kals, M., Mági, 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., Sneider, 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., Kuiviaeniemi, H., Tromp, G., Faucett, W.A., Li, R., Manolio, T.A., Sanderson, S.C., Kanny, 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., Pettova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delanoe, 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., Berner, 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., Fettersolf, S.N., Davis, F.D., Faucett, W.A., Kirchner, H.L., Mirshahi, U., Murray, M.E., 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, E.E., Murray, M.F., Overton, J.D., Hабеггер, L., Leader, J.B., Fettersolf, 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., Breeiling, 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-Zinzck, 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., Belbase, 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. Pharmacogenomic Personalized Med. 11, 59–66.

201. Tadaka, S., Hishinuma, E., Komaki, S., Motoike, I.N., Kawa, Wei, C.-Y., Yang, J.-H., Yeh, E.-C., Tsai, M.-F., Kao, H.-J., Lo, Aoki, Y., et al. (2021). Morp 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 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: scoping review. J Med. Inform. 7, e12239.

205. Shi, L., and Wang, Z. (2019). Computational strategies for scalable genomics analysis. Genes 10, 1017.

206. Chevrier, R., Foufi, V., Gaudet-Blavignac, C., Robert, A., and Lolis, C. (2019). Use and understanding of anonymization and de-identification in the biomedical literature: scoping review. J. Med. Internet Res. 21, e13484.

207. Manzoni, C., Kia, D.A., Vandrovcova, 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.

208. 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.

209. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Integrative approaches for large-scale transcrip-omics and proteome: the rise of omics data and their integration in biomedical sciences. Brief. Bioinform. 19, 1–22.

210. Manolio, T.A., Goodhand, P., and Ginsburg, G. (2020). ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? Int. J. Epidemiol. 42, 1–22.

211. Bowden, J., Pillig, L., Tadaka, S., Hishinuma, E., Hairad, 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.

212. 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.

213. Sheikhalishahi, S., Miotto, R., Dudley, J.T., Lavelli, A., Rinaldi, E., and Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: systematic review. JMR Med. Inform. 7, e12239.

214. Bowden, J., Pillig, 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., Caulde, 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. Runcharoen, C., Fukunaga, K., Sensorn, I., Iemwimangsa, N., Mpye, K., Matimba, A., Dzobo, K., Chirikure, S., Wonkam, A., Olowoyo, P., Ramsay, M., Skelton, M., and Stein, D.J. (2018). H3Africa: current perspectives. Pharmacogenomic Personalized Med. 11, 59–66.
et al. (2018). Genomic atlas of the human plasma proteome. Nature 558, 73–79.

228. Zhao, H., Rasheed, H., Nast, 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). Metabolic 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. (2019). Genomic Med. 12, 969.

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., Bierlein, E.L. (2020). Transfer learning enables prediction of CYP2D6 genotype across world populations. Genet. Med. 19, 26–31.

234. 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, eabf5637.

235. So, H.-C., Chau, C.-K.-L., Chiu, W.-T., Ho, K.-S., Lo, C.-P., Yim, S.H.-Y., and Shaw, P.-C. (2017). Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. Nat. Neurosci. 20, 1342–1349.

236. Subramanian, A., Narayan, R., Corseillo, 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.

237. Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Felix, E., Magiariños, M.P., Mosquera, J.F., Mutuwo, P., Nowotka, M., et al. (2019). ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 47, D930–D940.

238. Pushpakom, S., Iorio, F., Eyers, P.A., Escott, K.J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamme, C., et al. (2019). Drug repurposing: progress, challenges and recommendations. Nat. Rev. Drug Discov. 18, 41–58.

239. 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.

240. Nielsen, J. (2017). Systems biology of metabolism. Annu. Rev. Biochem. 86, 245–275.

241. 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.

242. Salvy, P., and Hatzimanikatis, V. (2020). The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. Nat. Commun. 11, 30.

243. Skime, M., 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.

244. 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.

245. 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. (2018). Global pharmacogenomics: the what and the why. Acta Biotheor. 69, 1437–1452.e17.

246. 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.

247. Knepper, T.C., and McLeod, H.L. (2018). When will clinical trials finally reflect diversity? Nature 557, 157–159.

248. Mills, M.C., and Rahal, C. (2019). A scientoometric review of genome-wide association studies. Commun. Biol. 2, 9.

249. Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The missing diversity in human genetic studies. Cell 177, 26–31.

250. 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.