

Genomewide Association Studies in Pharmacogenomics

Gregory McInnes¹, Sook Wah Yee², Yash Pershad³ and Russ B. Altman^{3,4,*}

The increasing availability of genotype data linked with information about drug-response phenotypes has enabled genomewide association studies (GWAS) that uncover genetic determinants of drug response. GWAS have discovered associations between genetic variants and both drug efficacy and adverse drug reactions. Despite these successes, the design of GWAS in pharmacogenomics (PGx) faces unique challenges. In this review, we analyze the last decade of GWAS in PGx. We review trends in publications over time, including the drugs and drug classes studied and the clinical phenotypes used. Several data sharing consortia have contributed substantially to the PGx GWAS literature. We anticipate increased focus on biobanks and highlight phenotypes that would best enable future PGx discoveries.

Genomewide association studies (GWAS) are an unbiased approach used to detect associations between genotype and phenotype in genotyping array or sequencing data. Since the first GWAS in 2005, GWAS have become a mainstay of genetics research.¹ GWAS have been performed for thousands of traits and have led to a better understanding of human genetics and the development of diagnostics and therapeutics.^{2,3} The growing abundance of genetic data linked with phenotype data has made GWAS an essential approach for understanding how genotype influences phenotype.

GWAS have been used to elucidate the mechanisms of interindividual differences in drug response.⁴ The study of the role genetics plays in drug response, broadly known as pharmacogenetics or pharmacogenomics (PGx; PGx is used interchangeably for pharmacogenetics and pharmacogenomics throughout this paper), has revealed clinically actionable insights that can improve patient outcomes, prevent severe adverse events, and reduce treatment costs.⁵ The earliest PGx GWAS were published between 2007 and 2008 around the same time.⁶ These studies demonstrated the unique power of GWAS and have provided important lessons for future studies of GWAS in PGx. For example, using only 74 cases and 130 controls, one study in 2007 identified associations between hepatic adverse events in patients on the oral direct thrombin inhibitor ximelagatran and the HLA locus.⁷ Since then, many more GWAS of the genetics of drug response have been published, studying dozens of drugs, identifying hundreds of genetic associations with drug response.

We performed a systematic review of the PGx GWAS literature and analyzed the number of studies over the last 13 years, which drugs and drug classes have been studied, and the ancestral populations of the study cohorts. We curated more than 400 papers to see which specific clinical end points have been used by researchers to study PGx and, based on these findings, make recommendations for data collection in biobanks to enable future PGx GWAS.

GENOMEWIDE ASSOCIATION STUDIES FOR PHARMACOGENOMICS

GWAS seek statistical associations between genomic loci and a phenotype of interest. An independent statistical test evaluates the potential association between a locus and the phenotype.⁸ The number of tests depends on the source of genetic data; genotyping arrays capture hundreds of thousands to millions of loci, whereas sequencing methods may identify millions of polymorphisms. Due to the enormous number of statistical tests, multiple hypothesis (e.g., Bonferroni) correction adjusts the resulting *P* values in order to limit false-positive results.⁹ The typical upper limit on *P* values for statistical significance for genotyping arrays is 5×10^{-8} .¹⁰ This cutoff is determined using a Bonferroni correction by dividing 0.05 (the frequently used *P* value significance threshold) by the total number of independent tests (i.e., the number of positions on the genotyping array) yielding the threshold that has been adjusted for the total number of statistical tests performed. To detect associations with this stringent *P* value cutoff, either the variants for the phenotype of interest must have large effect sizes or the study must be designed with large enough sample sizes to detect a modest effect size (for a detailed discussion of how sample size affects the ability to detect variants with different effect sizes, see Visscher *et al.*).^{2,11} Loci with a *P* value below the significance threshold become candidate variants for further investigation.

Effect sizes are also important for interpreting the impact of a variant. Whereas *P* values indicate a measure of confidence that the observed effect is a true association, effect sizes represent the magnitude of the likelihood that carrying a particular allele will lead to a change in phenotype. PGx variant effect sizes have been shown to be larger for dichotomous drug response traits than for other traits.¹² Because drug response phenotypes are typically not under the same selective pressure as variants that lead to disease, it is hypothesized that drug response alleles have larger effect sizes due to a lack of negative selective pressure. This makes PGx

¹Biomedical Informatics Training Program, Stanford University, Stanford, California, USA; ²Department of Bioengineering and Therapeutic Sciences, University of California at San Francisco, San Francisco, California, USA; ³Department of Bioengineering, Stanford University, Stanford, California, USA; ⁴Departments of Genetics, Medicine, Biomedical Data Science, Stanford, California, USA. *Correspondence: Russ B. Altman (russ.altman@stanford.edu)

Received April 2, 2021; accepted June 15, 2021. doi:10.1002/cpt.2349

variants especially clinically relevant because they more often lead to a change in phenotype.

With the release of the UK Biobank and other large phenotype-linked genetic datasets, GWAS have become a common analysis to run on many traits using hundreds of thousands of subjects.¹³ These studies benefit from massive cohorts of volunteer subjects with readily available phenotype data on many traits. Most of the phenotypes studied through these GWAS are easily defined, such as presence or absence of a disease or a quantitative measure, such as height.

PGx studies using GWAS face unique benefits and challenges when compared to disease genetics.⁴ Drug pharmacokinetics and pharmacodynamics operate within networks of proteins that are responsible for drug metabolism, transport, and the drug target. Genetic influences on drug response are frequently identified within these networks and are often either mono- or oligogenic with large effect sizes. For example, the association between *NUDT15* and thiopurine-induced leukopenia was detected using only 33 cases with a *P* value of 5×10^{-94} .¹⁴ Similarly, the earliest GWAS of warfarin maintenance dose included only 181 subjects to detect associations between dose and *CYP2C9* and *VKORC1*.¹⁵

Although small sample sizes sometimes suffice, PGx associations with more complex associations may be missed. Compared with disease genetics, collecting cohorts with well-phenotyped data to study PGx using GWAS is especially challenging. PGx studies necessitate a drug exposure in order to observe the phenotype of interest. Therefore, to collect a cohort, a researcher must first identify a population with the disease of interest, then treat those patients with the drug of interest, then observe interindividual differences in response. At each step, the potential patient population is winnowed down leaving study cohorts small, even when starting with massive, biobank-scale cohorts. One common approach used to alleviate this challenge is to use population controls rather than drug-matched controls. A study with population controls uses an existing cohort of samples as controls (e.g., the Wellcome Trust Case Control Consortium; <https://www.wtccc.org.uk/>) who may or may not have ever taken the query drug.¹⁶

THIRTEEN YEARS OF PGX GWAS

We performed a systematic review of the PGx GWAS literature.¹⁷ We identified PGx GWAS using two mechanisms: (1) entries in the GWAS Catalog who has studied phenotype was “Response to Drug,”³ and (2) literature curations from PharmGKB, which identify papers that used GWAS derive associations with drug response.¹⁸ This yielded 428 papers (as of March 27, 2021), which were then manually curated to verify accuracy. Of the 428 papers, 4 were removed for being unrelated to this study. A full list of all publications is available in **Table S1**.

We identified 424 PGx GWAS published between 2007 and 2020 in the GWAS Catalog (**Figure 1a**). PGx GWAS publications have been published in 123 journals, with *The Pharmacogenomics Journal* publishing the most. PGx GWAS represent 8.9% of all entries in the GWAS Catalog (**Figure 1b**). The highest percent of GWAS focusing on PGx was in 2015, with 17% of all GWAS being PGx related. The year 2015 also represents the year with the highest number of PGx GWAS published with 57 publications.

In the last 5 years, the median sample size of PGx GWAS is 1,220 (**Figure 1c**). The median sample size among PGx GWAS has remained mostly constant over the last 5 years (except for 2016), contrasted with all other GWAS, whose median sample size has grown steadily. The sample size in other forms of GWAS continues to increase, seeing a peak in 2019 with a median sample size of 10,584. However, we find that larger sample sizes do not necessarily yield more associations (**Figure S1**). We observe only a modest correlation between sample size and the number of significant associations discovered by a study ($R = 0.15$, Pearson's correlation between the base-10 logarithm of sample count and number of significant associations at P value $< 5 \times 10^{-8}$).

MOST STUDIED DRUGS

We found that 45 drug classes (Anatomical Therapeutic Chemical (ATC) Classification Level 3) have been studied using GWAS. These 45 drug classes include 8 out of the 12 top prescribed drug therapeutic category (<https://clinicalcalc.com/Downloads/Top250Drugs-DrugList.pdf>). Cancer drugs (ATC: L01) are the most studied, with 89 total publications, followed by antidepressants, antipsychotics, and lipid modifying agents. Fenofibrate is the most studied individual drug, with nine total publications performing GWAS on fenofibrate response. All PGx GWAS have yielded 586 total unique drug-variant associations (P value $< 5 \times 10^{-8}$; **Figure 2a**). We defined unique associations as a significant association identified in a PGx GWAS for a drug class that is not in linkage disequilibrium ($R^2 < 0.5$, calculated using LDlinkR using all populations in 1000 Genomes) with another significantly associated variant for the same drug class.^{19,20} The specific significance threshold varies between studies depending on the number of polymorphisms tested, but 5×10^{-8} is broadly accepted and used here for consistency.

We curated the total set of publications to better understand what specific phenotypes are used as clinical end points in PGx GWAS. We find that for most drug response GWAS, 56% of studies use therapeutic efficacy as the clinical end point, whereas 40% are related to adverse drug reactions (ADRs). We defined efficacy as any end point that directly studied differences in patient outcomes (e.g., recurrence-free survival). Whereas studies using ADRs as an end point use the incidence of an unwanted ADR as the phenotype of interest. We also identified three papers using a biomarker unrelated to the therapeutic mechanism and one paper that performed a GWAS to directly study drug metabolism.

The specific phenotypes measured vary by drug class, but several patterns emerged. For example, studies of response to asthma therapeutics nearly always study drug response and most often use the change in forced expiratory volume after treatment as the GWAS phenotype. Whereas studies of antidepressants or antipsychotics often use a quantitative measure of disease severity (e.g., the Hamilton Rating Scale for Depression) or incidence of a side effect (e.g., weight gain) as a clinical end point. There is no significant difference in the number of significant associations discovered by papers studying response or ADRs (**Figure 2b**). However, the sample size of GWAS of drug response is larger than those studying ADRs ($P = 0.01$, Student's *t*-test) with a median sample size of 738, compared to 669 (**Figure S2**).

Pharmacogenomics GWAS publication statistics

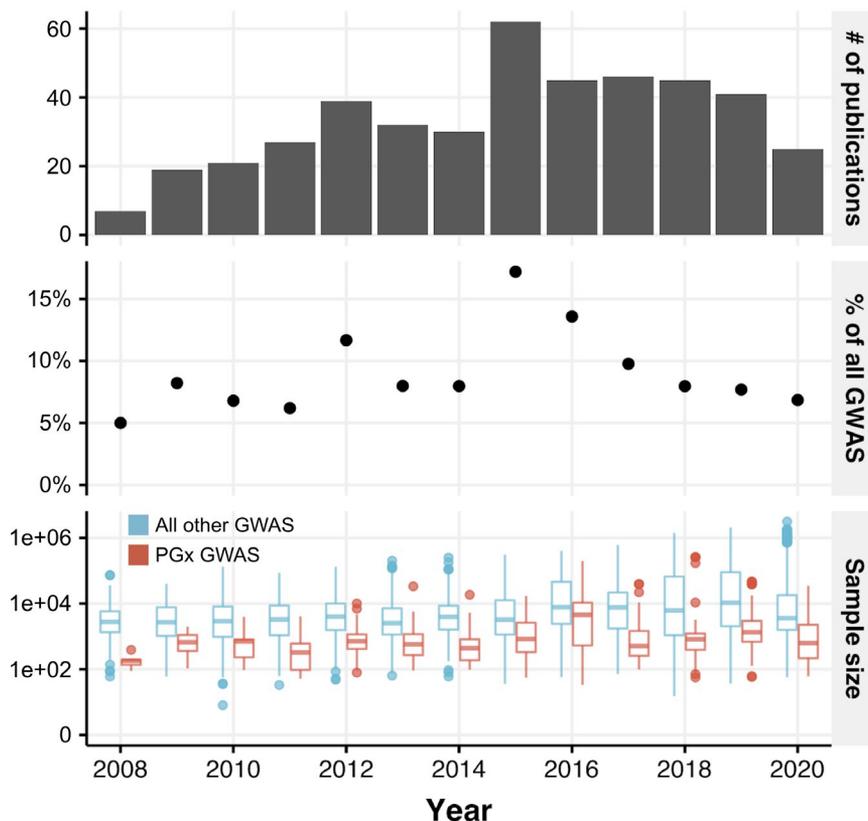


FIGURE 1 Statistics of pharmacogenomics (PGx) genomewide association studies (GWAS) performed from 2008 through June 2020. (a) The total number of publications studying PGx each year. (b) The percent of all GWAS published that study PGx in the GWAS Catalog. (c) The cohort size of PGx GWAS (blue) compared to the cohort size of all other GWAS (red), derived from the GWAS Catalog. The year of publication is derived from PubMed which may differ from the actual publication date. Publications were queried from GWAS Catalog on March 27, 2021.

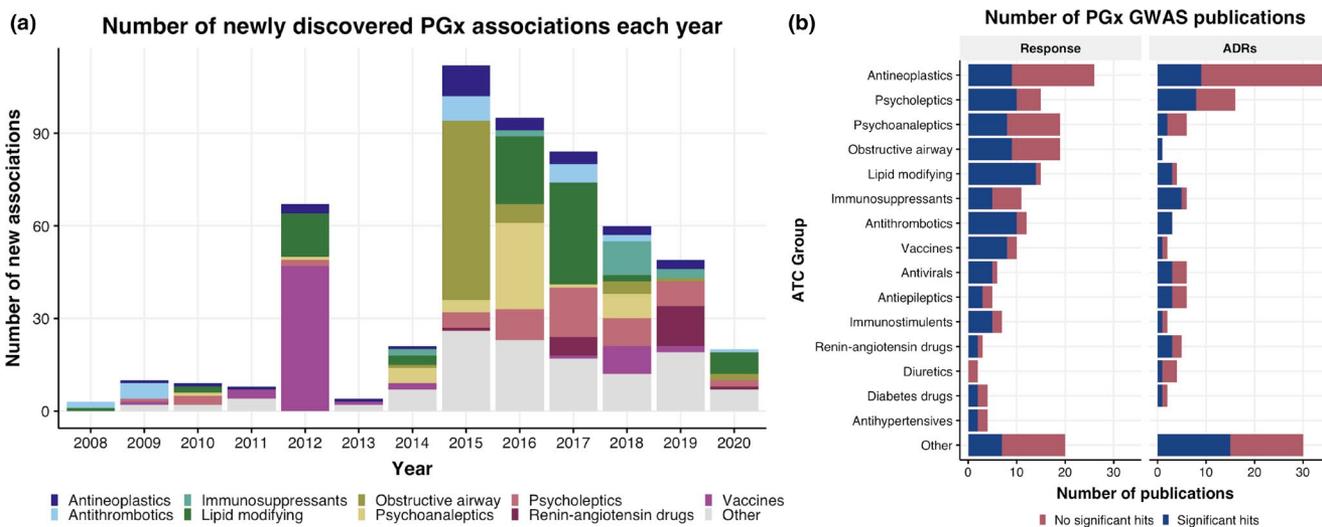


FIGURE 2 (a) Number of newly discovered drug response associations each year that had not previously been identified ($P < 5 \times 10^{-8}$) or in linkage disequilibrium with a previously identified variant ($R^2 < 0.5$). Colors represent Anatomical Therapeutic Chemical (ATC) groups. Only the top nine ATC groups ranked by number of unique associations are shown. All other ATC groups are grouped into “Other.” (b) The sixteen most studied ATC groups using genomewide association studies (GWAS) and whether response or adverse drug reactions (ADRs) were the focus of the study. Colors represent whether the study had significant findings ($p < 5 \times 10^{-8}$). All ATC groups not in the top 16 are grouped into “Other.” The large number of vaccine associations discovered in 2012 were derived mostly from a single publication studying side effects of the smallpox vaccine.¹⁰² PGx, pharmacogenomics.

Cancer drugs

Antineoplastic agents, drugs used to treat cancer, are the most frequently studied drug class using PGx GWAS. These drugs are of particular interest because the indication is severe and high toxicity of antineoplastics leads to many ADRs. The great interest in improving outcomes for patients with cancer broadly in the medical field is reflected in the abundance of PGx GWAS as well. This section focuses on cancer PGx studies.

We identified 94 GWAS of cancer drugs. We find that 61 of these studies sought to study PGx influence on outcomes of individual drugs, whereas the rest focused on broader drug classes or combination therapies. The most studied individual cancer drugs are methotrexate and paclitaxel, with six studies each. Six independent studies have investigated PGx of paclitaxel response; four of which studied the genomics of paclitaxel-induced peripheral neuropathy and have implicated the gene *SIPRI* in peripheral neuropathy risk.²¹ Other frequently studied drugs are cisplatin and other platinum compounds (13 studies), and anthracyclines (5 studies).

The associations investigated by cancer PGx studies are divided between response and ADRs. Among these 59 GWAS, 33 looked for genetic associations with ADRs, 25 studied drug response, and one studied genetic influence on a related biomarker. The most frequently studied phenotype is survival. Twelve studies evaluated heterogeneity in drug response by performing a GWAS, regressing on the amount of time post-treatment patients survived. The most frequently studied ADRs are drug-induced peripheral neuropathy (8 studies) and drug-induced agranulocytosis (4 studies).

Despite the great interest in the PGx of cancer treatment, most PGx GWAS do not find significant associations with response to treatment. Only 17 of the 59 studies find any significant association with the studied phenotype (P value $< 5 \times 10^{-8}$). This may be due to other confounding factors that weakened the associations, such as disease heterogeneity and prior therapies.²²

PHARMACOGENES IN GWAS

Years of PGx research have led to curated lists of important pharmacogenes, which are known to modulate drug response often by being involved in drug metabolism or transport. To determine how many of these known important pharmacogenes are among the top loci in PGx GWAS, we combined lists of genes from PharmGKB (<https://www.pharmgkb.org/vips>), the US Food and Drug Administration (FDA) Table for Pharmacogenomics Biomarkers (<https://www.fda.gov/medical-devices/precision-medicine/table-pharmacogenetic-associations>), and the Clinical Pharmacogenetic Implementation Consortium (CPIC; <https://cpicpgx.org/>) into a final list of 210 genes (Table S4). We then queried variants in the gene loci (plus 50 kilobases upstream and downstream) from the reported associations in the GWAS Catalog for our list of PGx GWAS. Figure 3 shows 46 drugs that have at least one significant association with one of the pharmacogenes. We identified 45 drug-gene pairs with significant associations (P value $< 5 \times 10^{-8}$). In this section, we describe several noteworthy findings from these PGx GWAS in relation to important pharmacogenes.

GWAS confirm findings from candidate gene studies

The very nature of GWAS enables interrogation of wide swaths of the genome, empowering a much broader search for associations than candidate gene studies. Previous candidate gene studies have identified polymorphisms in genes associated with drug response or hypersensitivity and PGx GWAS have confirmed many of these associations. For example, GWAS have confirmed *CYP2C9* and *VKORC1* as being strongly associated with warfarin maintenance dose,^{15,23–27} which was previously known through candidate gene studies.^{28,29} GWAS also revealed novel associations of *CYP4F2* with warfarin dose by controlling for the strong effects of *CYP2C9* and *VKORC1*.²³

The HLA region

Prior to PGx GWAS, genetic polymorphisms in *HLA* locus have been shown to play a role in a broad range of drug hypersensitivity or rare toxicity.³⁰ PGx GWAS confirmed those previously known associations with the *HLA* locus and reported new associations, such as sulfasalazine-induced agranulocytosis.³¹ The large effect sizes for associations of *HLA* alleles with drug-induced hypersensitivity or rare toxicity enable new discoveries even in studies with small sample sizes and consequently only a few cases. For instance, a study with only 30 cases identified an association between *HLA-B* and sulfasalazine-induced agranulocytosis. Interestingly, PGx GWAS revealed that polymorphisms in *HLA* locus could also be important for hepatitis B vaccine response as well as interferon-beta therapy.^{32–34}

Discovery in non-European cohorts

GWAS cohorts have historically been over-represented by subjects of European descent (discussed in detail in the next section).³⁵ Allele frequencies in pharmacogenes can vary greatly across global populations and can lead to heterogeneity in drug response.^{36–38} This allelic variation in pharmacogenes leads to an opportunity to discover novel drug-gene associations in non-European populations. For example, thiopurine toxicity is known to be associated with variants in *TPMT*; however, using a non-Europeans cohort, a novel locus, *NUDT15*, was identified.^{14,39} The causal variant in *NUDT15* is most common in East Asians and Hispanics, but rare in Europeans. Based on the results of these studies and subsequent replications, the FDA approved labels state that testing for *TPMT* and *NUDT15* deficiency should be considered when prescribing thiopurine drugs.⁴⁰ Other population-specific PGx GWAS in individuals of African descent have led to discoveries of ethnic specific variant associations with warfarin dose.^{25,41–43} For example, a study by De *et al.*, which have led to discovery of an ethnic specific variant upstream of a biological relevant gene, *EPHA7*, with population-specific warfarin-associated bleeding.⁴³

Novel substrate discovery

PGx GWAS has successfully identified novel substrates for transporters. For example, GWAS provided the first evidence that allopurinol is a substrate of ABCG2 through GWAS.⁴⁴

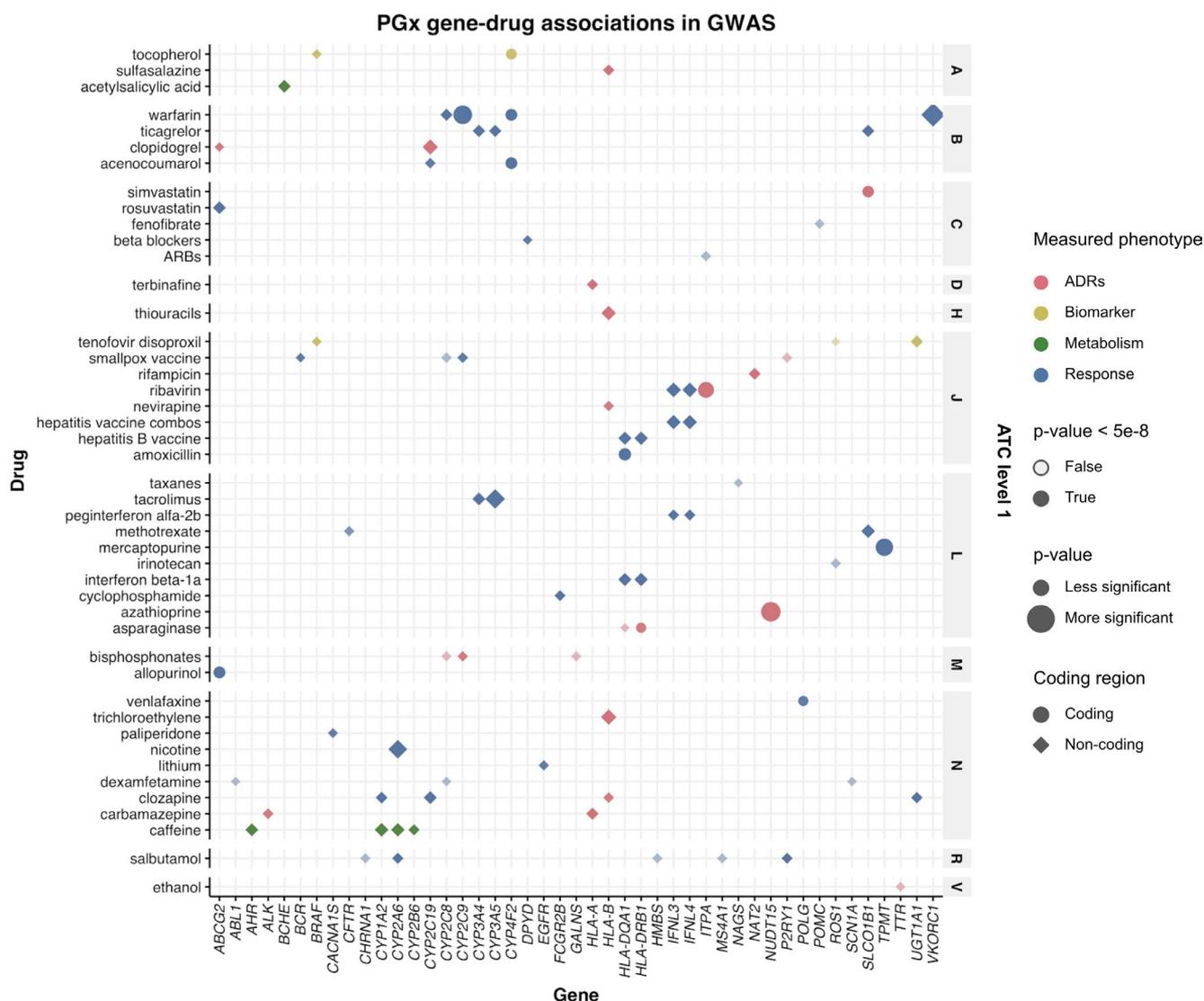


FIGURE 3 Gene-drug associations identified using genomewide association studies (GWAS). Associations included in this figure were identified through the reported associations in GWAS Catalog. Each point represents a reported association between a gene region, plus 50 kilobases upstream and downstream, and a drug response measured phenotype. The variant with the lowest P value within the locus from any study was selected. The most specific drug or drug class was selected based on Anatomical Therapeutic Chemical (ATC) code. For example, studies of general statin use are not included because there are studies specifically for rosuvastatin and simvastatin. The letters on the right side of the figure represent ATC level 1 code of the drug's ATC code. Absence of a dot means that there was no study that reported an association for that drug-gene pair for that specific drug in GWAS Catalog. Only drugs and genes with at least one association are shown. Neighboring genes may share associations if they are within 50 kilobases. Circles indicate variants that are reported by GWAS Catalog to be in coding regions (e.g., missense variants), diamonds indicate variants in noncoding regions (e.g., intronic). For example, warfarin: single-nucleotide polymorphisms in noncoding region of *CYP2C8* (diamond), coding region of *CYP2C9* (circle), coding region of *CYP4F2* (circle), and noncoding region of *VKORC1* (diamond), are significantly associated with warfarin response (blue color diamond or circle) at $P < 5 \times 10^{-8}$. PGx, pharmacogenomics.

Pleiotropy between disease and pharmacogenomic traits

Among these significant associations between pharmacogenes and PGx traits, two of the genes are also associated with disease traits relevant to the drug treatment: (1) *ABCG2* and its association with allopurinol drug response, along with serum uric acid levels and gout,⁴⁵ and (2) *IFNL3/IFNL4* locus encoding interleukin 28B (IL28B) and its association with hepatitis C infection and clearance^{46–48} and response to peg-interferon therapy for hepatitis C.⁴⁹ Other examples beyond pharmacogenes that are worth

mentioning are statin response with single nucleotide variants in genes^{50–52} that affect lipid LDL-levels (e.g., *LPA*, *APOE*, and *PCSK9*),⁵³ as well as metformin response with single-nucleotide polymorphism in glucose transporter, *SLC2A2*,⁵⁴ that also affect plasma glucose and HbA1c levels.⁵⁵

STUDIED POPULATIONS

We next studied the distribution of ethnicities in PGx GWAS participants by extracting the broad ancestral population reported

in the GWAS Catalog for each study accession and mapping those to global populations.⁵⁶ Individuals of European descent comprise the vast majority of PGx GWAS participants. We find that 88% of study participants are of European ancestry in the discovery phase of PGx GWAS, with the next highest studied population being individuals of Hispanic/Latin American descent at 4% (**Figure 4d**). In the replication phase of experiments, Europeans represent 73% of study subjects, with a greater number of non-European individuals contributing to replication experiments than in the discovery phase (**Figure 4e**). Although the sample size of PGx GWAS has grown since 2008, the proportion of non-European individuals of non-European descent included in PGx GWAS has remained largely unchanged since 2008 (**Figure 4a,b**). There is an increasing frequency of very large cohorts in PGx GWAS, however, these studies primarily comprise European subjects. These studies with very large sample sizes resulted from the use of an alternative approach to conduct PGx GWAS, where self-reported questionnaires from 23andMe survey data were used to assess antidepressant efficacy and side effects.⁵⁷ A total of 48,000 research participants answered surveys related to antidepressant used and health history. This largest PGx GWAS analysis also includes 190,000 healthy controls free of known neuropsychiatric diseases.

Discovery in non-European cohorts

GWAS cohorts have historically been over-represented by subjects of European descent.³⁵ Although Europeans account for the vast

majority of studied subjects, we find that, at the individual study level, there is moderately better representation of broader demographics. In total, 53% of all PGx studies focus solely on individuals of European descent, with a median sample size of 4,300. The next most studied population is Asians, comprising 16% of all studies, followed by Africans at 15%.

IMPACTFUL CONSORTIA AND COHORTS IN PGX GWAS

Data from patients recruited for other studies or clinical trials are sometimes useful for PGx analyses. This is especially true in GWAS where large sample sizes are needed. Once phenotypic and genotypic data are collected, they can be used repeatedly for new discovery cohorts independently or included in meta-analyses. The data can also be used as a control cohort for a separate study of a different phenotype. Additionally, consortia can wield funding and resources to enable research. Consortia can bring together scientists across institutions and disciplines forming collaborations that lead to studies that may not have otherwise been possible.

We curated the manuscripts in our set to determine which cohorts or consortia were the most impactful throughout PGx GWAS. We counted the number of times specific cohorts were used as either a discovery or replication cohort in any GWAS, as well as the number of GWAS consortia produced. We find that many cohorts are used for various studies, including disease associations as well as PGx research. In **Figure 5** we show the 17

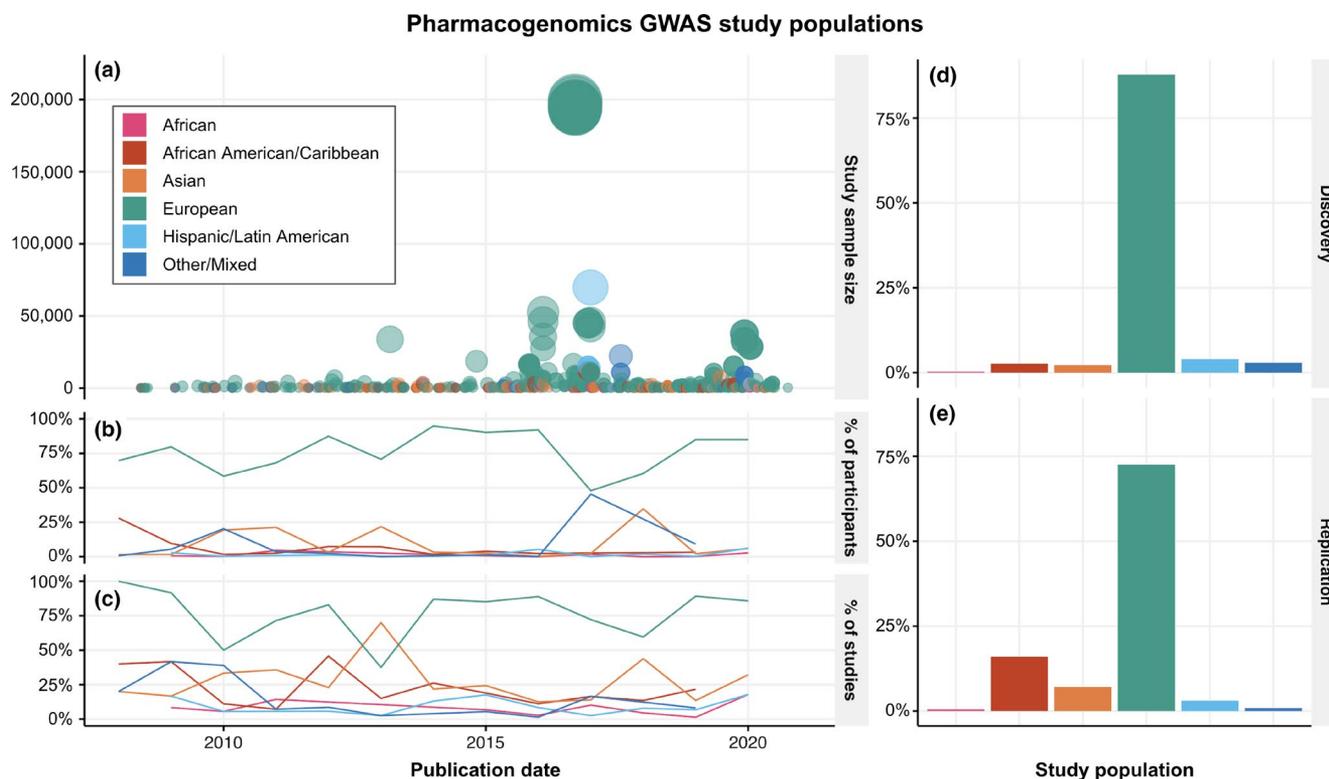


FIGURE 4 Pharmacogenomics (PGx) genomewide association studies (GWAS) populations from 2008 to 2020 show European bias in study participants. Each color represents an ancestral population. (a) Discovery cohort size for PGx GWAS over time. Dot size is correlated with study size. (b) Percentage of total PGx GWAS participants in discovery cohorts belonging to each ancestral population over time. (c) The percentage of PGx GWAS focusing on each ancestral population over time. (d) Percent of all PGx GWAS participants in discovery cohorts based on their ancestral population. (e) Percent of all PGx GWAS participants in replication cohorts based on their ancestral population.

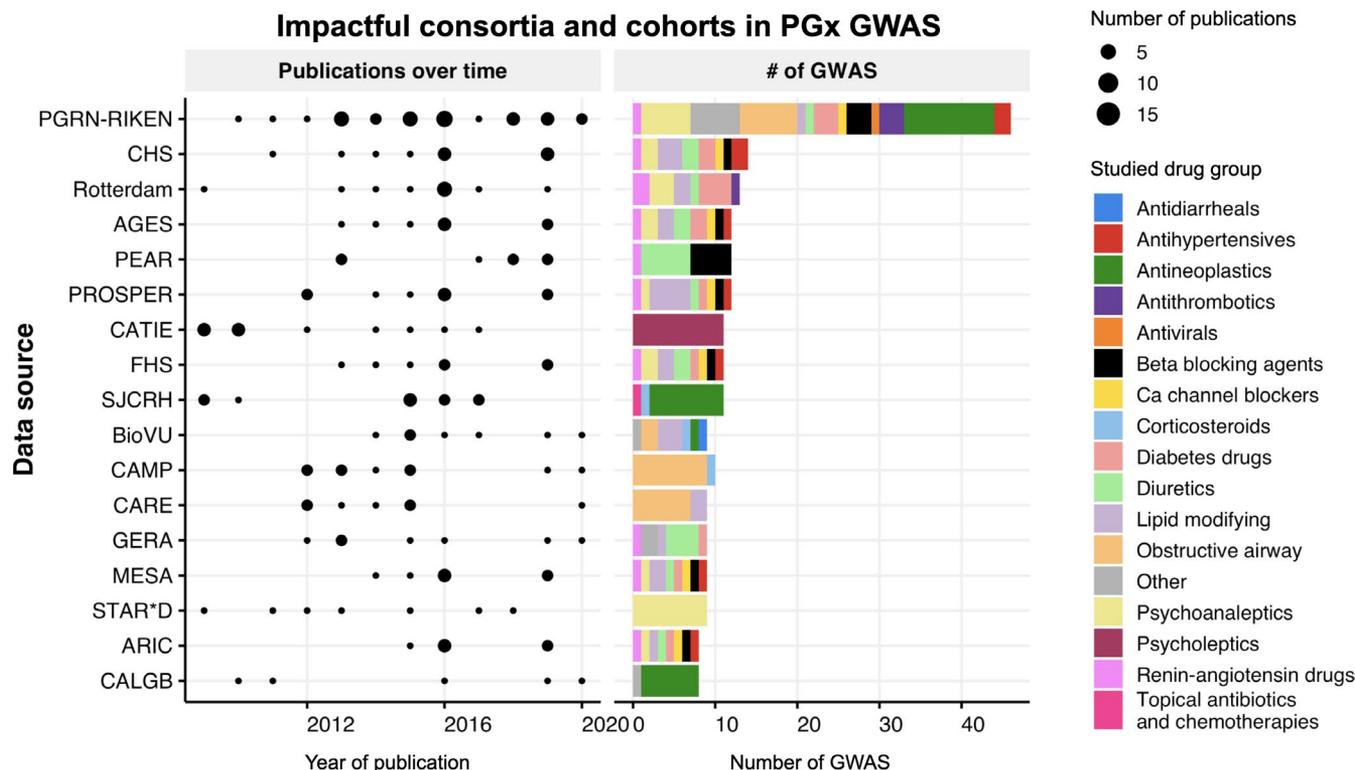


FIGURE 5 Impactful consortia and cohorts in pharmacogenomics (PGx) genomewide association studies (GWAS). Each row represents a single cohort, consortia, or institution, and its length of the bar in the right-most part of the figure represents the number of published GWAS data from that cohort has contributed to (a single publication may contain more than one GWAS). Colors represent drug Anatomical Therapeutic Chemical (ATC) groups. Dots on the left side of the figure represent the annual publication frequency of each consortium or cohort. Larger dots indicate more publications. Abbreviations: CHS: Cardiovascular Health Study; Rotterdam: Rotterdam studies; AGES: Age, Gene, Environment, Susceptibility; PEAR: Pharmacogenomic Evaluation of Antihypertensive Responses; PROSPER: Prospective study of Pravastatin in the Elderly at Risk; CATIE: Clinical Antipsychotic Trials of Intervention Effectiveness; FHS: Framingham Heart Study; SJCRH: St. Jude Children's Research Hospital; BioVU: Vanderbilt University Biobank; CAMP: Childhood Asthma Management Program; CARE: Childhood Asthma Research and Education; GERA: Genetic Epidemiology Research on Adult Health and Aging; MESA: Multi-Ethnic Study of Atherosclerosis; STAR*D: Sequenced Treatment Alternatives to Relieve Depression; ARIC: Atherosclerosis Risk in Communities; CALGB: Cancer and Leukemia Group B.

consortia and cohorts that contributed most to PGx GWAS, as well as the time period where papers were published using the cohorts' data.

Consortia focusing on PGx to advance research have been established with a goal of combining cohorts collected from different investigators for PGx GWAS. For example, the International Serious Adverse Events Consortium, International Clopidogrel Pharmacogenomics Consortium, International drug-induced liver injury (DILI) consortium, and many others outside of North America, such as Japan PGx Data Science Consortium and African Pharmacogenomics Consortium.⁵⁸ These have led to several discoveries of new genomewide significant loci for PGx studies, such as the discovery of single-nucleotide polymorphisms in *SLC2A2*, a glucose transporter, as determinant of interindividual differences in response to anti-diabetic drug, metformin, by the MetGen Consortium,⁵⁴ the discovery of new loci in addition to *CYP2C19* for clopidogrel response by the International Clopidogrel Pharmacogenomics Consortium,⁵⁹ and the discovery of *PTPN22*, a new loci beyond the HLA-locus for drug-induced liver injury by the International DILI consortium.⁶⁰ Many of the studies identified are meta-analyses, combining data from many smaller studies

into larger GWAS that are better powered to detect associations with small effect sizes.⁶¹

A unique international collaborative effort worth highlighting that has led to more than 45 published GWAS in PGx studies is the PGRN-RIKEN Global Alliance (**Table S2**). This collaboration was founded in 2008, under the leadership of Yusuke Nakamura from The University of Tokyo, along with Kathleen Giacomini and Mark Ratain, two well-established PGRN NIH-funded investigators. From 2008 to 2016, RIKEN (Center for Genomics Medicine and Center for Integrative Medical Sciences) provided full support for genomewide genotyping to investigators from Pharmacogenomics Research Network (PGRN), who had well-characterized cohorts with PGx phenotype along with collected DNA samples. This unique collaboration supported 46 distinct GWAS with more than 56,000 multi-ethnic DNA samples, which were genotyped and analyzed by the alliance (<http://pgrn2016.weebly.com/riken-projects.html>).^{62,63} The individual studies included samples from various sources, including clinical trial cooperative groups (e.g. CALGB⁶⁴), PGx related consortia (e.g., International Warfarin PGx Consortium, International Clopidogrel PGx Consortium), and electronic health-record linked with biorepository of DNA (e.g., BioVU and

RPGEH). The collaborations have led to the first PGx studies for various drug classes, for example, response, adverse drug response, and drug levels of aromatase inhibitors in patients with breast cancer,^{65–67} paclitaxel-induced peripheral neuropathy,⁶⁸ response to various drugs to treat asthma.^{69,70} Although the PGRN-RIKEN sunset in 2016, the collected data have led to continuous publications by the investigators ranging from functional genomic studies to meta-analysis and polygenic risk score analysis (Table S3). The summary statistics from several of the PGx GWAS by the alliance are disseminated and populated here (<https://www.pgrn.org/riken-gwas-statistics.html>). Overall, this collaboration shows the powerful impact of PGRN-RIKEN in PGx GWAS, which serves as a model in accelerating the PGx GWAS.

FUTURE PERSPECTIVES

Although sample sizes have increased dramatically, the number of novel associations detected has not (Figure 2a). In order to detect existing unexplained heritability in drug response, the field will need larger sample sizes, more diverse cohorts, and a broader array of statistical tests. In this section, we describe future perspectives that may help to further develop PGx association studies.

Biobanks

The growing number and availability of biobanks providing phenotype-linked genotype data offers an opportunity. Huge amounts of genotype data linked with clinical data offers an opportunity to detect associations at an unprecedented scale. Resources like the BioBank Japan (BBJ)⁷¹ and UK Biobank have already generated extreme interest and huge numbers of discoveries, but to be useful for PGx, these resources must have sufficient drug data. Most of the studies we identified were drawn from cohort studies specifically recruited to study PGx. However, with the availability of biobanks, such as UK Biobank and the coming release of data from the All of Us Research Program, studying there is an opportunity to study drug response at unprecedented scales. Furthermore, there are enormous genomic data initiatives occurring globally, working diligently to integrate genomics into healthcare, such as the 100 million genomes through the Chinese Precision Medicine Initiative.⁷² These projects offer rich phenotype data linked to genotype data, but in order to study PGx sufficient phenotype data needs to be available.

Several key features are important for conducting retrospective PGx-focused cohort studies using biobanks. At a minimum, information about patient demographics (e.g., age and sex), the name of the prescribed drug, the date of the prescription, and diagnosis codes for any subsequent encounters with the medical system following the initial prescription. Ideally, the drug formulation would be represented using a standardized terminology that maps the active ingredients of the drug, route of administration, and dose to a representative code, such as the systematized nomenclature of medicine clinical terms (SNOMED CT) or the National Health System's dictionary of medicines and devices (dm+d). This core set of phenotype data would enable many of the PGx GWAS of PGx of side effect incidence to be replicated, assuming sufficient samples exist. These data are already available in the UK Biobank

for 230,000 participants in the form of longitudinal clinical data derived from general practitioner visits.⁷³

Response studies may require additional data, which could be cumbersome to collect. For example, we find that many efforts studying the genetics of heterogeneity in depression treatment use an instrument, such as the Hamilton Rating Scale for Depression (HAM-D), before and after treatment as a quantitative measure of the change in phenotype.⁷⁴ These tests may not be regularly conducted prior to and following treatment, and therefore would not be available in biobanks collecting the data from providers. This presents a challenge to researchers studying the PGx of depression to design a phenotype based on the phenotypes that are available, rather than an ideal quantitative measure. Periodic self-reported phenotypes may be sufficient for detecting associations in treatment response, provided that they are done before and after treatment.⁷⁵

We find that quantitative measures of biomarkers are frequently used to measure response and toxicity risk and study PGx effects. For example, biomarkers, such as blood lipid levels, are frequently used to measure statin response, which has known genetic influences. Biobanks already collect measurements of key biomarkers and researchers have conducted GWAS to study the genetic influences of interindividual differences in these levels.⁷⁶ Combining periodic measurements of biomarkers with recent drug exposures could allow for PGx focused studies of the change in biomarkers in response to treatment.

In instances where longitudinal clinical data are limited and only information about patient prescriptions is available, it may still be possible to conduct PGx GWAS. A recent study found that by performing a GWAS on which type of statin (e.g., atorvastatin vs. simvastatin) a subject was prescribed at the time of the UK Biobank participant intake survey recapitulated previously identified statin response alleles.⁷⁷ Performing GWAS on drug selection may discover associations with ADRs or response that have inadvertently led the patient to take one drug over another.

Diversity in PGx Studies

There is a known bias toward the inclusion of Europeans in genetics research,⁷⁸ and here we show that PGx suffers from the same issue. Diversification of study populations in PGx studies may lead to the discovery of more associations and thus yield greater clinical outcomes.³⁵ Pharmacogenes are under lower evolutionary constraint than disease genes,⁷⁹ which leads to vastly different allele frequencies between global populations.³⁷ Additionally, non-European populations harbor a higher frequency of previously unseen rare deleterious variants in pharmacogenes, likely as the result of being understudied.³⁷ Narrowly focusing research on a single population limits the impact of PGx by limiting the degree to which important associations can be discovered.

GWAS must also move beyond studying uniform ethnic populations. Much of the global population admixed. Rather than seeking to avoid admixture, methods need to be developed to account for admixture and multi-ethnic cohorts. Trans-ethnic GWAS have been performed successfully, but remain the exception rather than the rule.⁸⁰ It is already standard procedure to include values derived from principal component analysis as covariates in the GWAS to account for some population diversity, we must build upon this

strategy to expand the reach of PGx GWAS. Methods have been developed to determine local ancestry, which enables the inclusion of admixed individuals in GWAS, boosting power to detect associations.⁸¹ Building upon these methods will be critical for furthering our understanding of PGx.

Beyond GWAS

With the increasing availability of sequencing data (as opposed to genotyping data) it will be possible to perform association tests that account for rare variants. Nearly all the studies identified through this review use genotyping data, which has limited ability to identify rare variants, even when probes are designed to detect them.⁸² GWAS may not detect associations between independent rare variants unless the effect size is extremely large, but tests that aggregate the effects of rare or deleterious variants in a region or gene can be used to detect an effect.⁸³ Such tests can be used to detect PGx effects of rare variants in sequencing and exome data and have been shown to be able to detect the influence of deleterious variants in *CYP2D6* on ADRs related to opioid use.⁸⁴

Phenome-wide association studies (PheWAS) have grown in interest as a method of identifying the effect of individual variants across a broad range of phenotypes.⁸⁵ PheWAS, much like GWAS, perform independent statistical tests analyzing the association between genotype and phenotype. Rather than look across the genome for genetic associations with a phenotype, PheWAS look across a range of phenotypes, or the phenome, for phenotype associations with an individual genotype. Using PheWAS, it will be possible to study the effect of individual variants or PGx phenotypes (e.g., *CYP2D6* metabolizer status) across a range of drug response phenotypes. Recent work showed that there are significant associations between cytochrome P450 phenotypes and maintenance dose in the UK Biobank, demonstrating that PGx studies that more broadly link PGx phenotypes and drug response phenotypes may enable further discovery.⁸⁶

Polygenic risk scores (PRS) are a growing area of interest in disease genetics and may have applications in predicting individual drug response. PRS generates scores for an individual based on the sum of a large number of variants with small effect sizes throughout the genome initially identified through GWAS. These scores can be used to predict the probability of disease occurrence, or possibly drug response. Lanfear *et al.* retrospectively demonstrated that PRS may identify patients with heart failure who have increased survival benefit when treated with beta blockers.⁸⁷

With greater availability of sequencing data, studies focused on the effects of structural variants, including copy number variants, will be feasible. There are several known pharmacogenes with frequent structural variation that have a strong influence on drug response, including *CYP2D6*.⁸⁸ Genome sequencing data are best equipped for the identification of structural variants, and studies of the more than 50,000 TOPMED genomes has shown that there is much more heterogeneity in *CYP2D6* structural variation than had been previously shown.⁸⁹ The ability to accurately detect these events will improve the ability to study associations with drug response. It will be possible to perform PheWAS of drug response to study the broader effect of structural variants on drug response.⁹⁰

Functional characterization

Other considerations to complement PGx GWAS are associations of extreme PGx phenotypes and functional genomic studies. As noted, large effect size could be achieved with extreme traits such as those from rare drug toxicities.⁷ However, characterization of such extreme response phenotypes has been slow due to small sample sizes and challenges in identifying them.⁹¹ Future studies are needed to know whether extreme drug response traits could be effective ways to uncover novel loci. PGx GWAS have revealed plausible mechanisms underlying drug response or toxicities. Despite this, there are challenges in follow-up studies from GWAS and these have been reviewed recently.⁹² Investigators have applied various technologies, tool sets, and complex analyses to unravel the variants discovered, such as those examples from PGRN-RIKEN collaboration (see **Table S2**). Two examples worth mentioning about followed up from initial PGx GWAS are the used of multiple functional genomics studies to discover additional mechanism of drug action of anastrozole⁹³ and the utilization of massively parallel variant function assays to determine 3,000 missense variants in *NUDT15*⁹⁴ and more than 6,000 missense variants in *CYP2C9*.⁹⁵ This increase in functional data enables computational prediction of variant function using approaches, such as machine learning, which may someday increase the clinical utility of rare variants as they are detected in patients.^{96–98}

CONCLUSION

Associations derived from GWAS only serve as a starting place for understanding the influence of genetics on drug response.⁹⁹ The discovered associations are frequently not causal, but rather in linkage disequilibrium with the causal variant. Fine mapping must be performed to identify causal variants such that they can be used for diagnostic purposes.¹⁰⁰ Even more importantly, findings must be reproducible by subsequent studies in external cohorts to confirm associations. Then, organizations, such as the CPIC, develop therapeutic guidelines that provide clinical recommendations based on a patient's genotype. Many such guidelines have been developed and many of the drug-gene associations represented in those guidelines have evidence of an association observed through GWAS.¹⁰¹ Bringing these discoveries to their full clinical utility is critical to meet the goal of bringing PGx into the clinic.

The field of PGx has greatly expanded the collective knowledge of drug response genetics through the use of GWAS. Hundreds of associations between genes and drugs have been discovered and, as large datasets become increasingly available, many more will be discovered. There are opportunities in therapeutic areas not yet studied. Notably, common drugs used in these therapeutic categories, such as ophthalmic and otolaryngological drugs, dermatologic drugs, renal drugs and gastrointestinal drugs, have not been extensively studied in GWAS.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Kathy Giacomini for helpful comments on the manuscript and discussions about the PGRN-RIKEN collaboration.

FUNDING

G.M. is supported by the Big Data to Knowledge (BD2K) from the National Institutes of Health (T32LM012409). S.W.Y. is supported by NIH/National Institute of General Medicine Sciences (R01GM117163). R.B.A. is supported by NIH/National Institute of General Medical Sciences PharmGKB resource (U24HG010615), NIH GM102365, and the Chan Zuckerberg Biohub.

CONFLICT OF INTEREST

R.B.A. is a stockholder in Personalis.com, 23andMe.com. All other authors declared no competing interests for this work.

DATA AVAILABILITY STATEMENT

All code used for this analysis is freely available on GitHub (<https://github.com/gregmcinnis/PGx-GWAS-Review>).

© 2021 The Authors. *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

1. Klein, R.J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 (2005).
2. Visscher, P.M. *et al.* 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
3. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
4. Motsinger-Reif, A.A. *et al.* Genome-wide association studies in pharmacogenomics: successes and lessons. *Pharmacogenet. Genomics* **23**, 383–394 (2013).
5. Krebs, K. & Milani, L. Translating pharmacogenomics into clinical decisions: do not let the perfect be the enemy of the good. *Hum. Genomics* **13**, 39 (2019).
6. Crowley, J.J., Sullivan, P.F. & McLeod, H.L. Pharmacogenomic genome-wide association studies: lessons learned thus far. *Pharmacogenomics* **10**, 161–163 (2009).
7. Kindmark, A. *et al.* Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J.* **8**, 186–195 (2008).
8. Cantor, R.M., Lange, K. & Sinsheimer, J.S. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.* **86**, 6–22 (2010).
9. Johnson, R.C. *et al.* Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genom.* **11**, 724 (2010).
10. Jannot, A.-S., Ehret, G. & Perneger, T. $P < 5 \times 10^{-8}$ has emerged as a standard of statistical significance for genome-wide association studies. *J. Clin. Epidemiol.* **68**, 460–465 (2015).
11. Hong, E.P. & Park, J.W. Sample size and statistical power calculation in genetic association studies. *Genomics Inform.* **10**, 117–122 (2012).
12. Maranville, J.C. & Cox, N.J. Pharmacogenomic variants have larger effect sizes than genetic variants associated with other dichotomous complex traits. *Pharmacogenomics J.* **16**, 388–392 (2016).
13. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine* **12**, e1001779 (2015).
14. Yang, S.-K. *et al.* A common missense variant in NUDT15 confers susceptibility to thiopurine-induced leukopenia. *Nat. Genet.* **46**, 1017–1020 (2014).
15. Cooper, G.M. *et al.* A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* **112**, 1022–1027 (2008).
16. Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
17. Mills, M.C. & Rahal, C. A scientometric review of genome-wide association studies. *Commun. Biol.* **2**, 9 (2019).
18. Whirl-Carrillo, M. *et al.* Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* **92**, 414–417 (2012).
19. Myers, T.A., Chanock, S.J. & Machiela, M.J. LDlinkR: An R Package for rapidly calculating linkage disequilibrium statistics in diverse populations. *Front. Genet.* **11**, 157 (2020).
20. Consortium, Genomes Project *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
21. Chua, K.C. *et al.* Genomewide meta-analysis validates a role for S1PR1 in microtubule targeting agent-induced sensory peripheral neuropathy. *Clin. Pharmacol. Ther.* **108**, 625–634 (2020).
22. Wheeler, H.E., Maitland, M.L., Dolan, M.E., Cox, N.J. & Ratain, M.J. Cancer pharmacogenomics: strategies and challenges. *Nat. Rev. Genet.* **14**, 23–34 (2013).
23. Takeuchi, F. *et al.* A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet.* **5**, e1000433 (2009).
24. Cha, P.-C. *et al.* Genome-wide association study identifies genetic determinants of warfarin responsiveness for Japanese. *Hum. Mol. Genet.* **19**, 4735–4744 (2010).
25. Perera, M.A. *et al.* Genetic variants associated with warfarin dose in African-American individuals: a genome-wide association study. *Lancet* **382**, 790–796 (2013).
26. Parra, E.J. *et al.* Genome-wide association study of warfarin maintenance dose in a Brazilian sample. *Pharmacogenomics* **16**, 1253–1263 (2015).
27. Liu, N. *et al.* Influence of common and rare genetic variation on warfarin dose among African-Americans and European-Americans using the exome array. *Pharmacogenomics* **18**, 1059–1073 (2017).
28. Higashi, M.K. *et al.* Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. *JAMA* **287**, 1690–1698 (2002).
29. Wadelius, M. *et al.* Association of warfarin dose with genes involved in its action and metabolism. *Hum. Genet.* **121**, 23–34 (2007).
30. Daly, A.K. Pharmacogenomics of adverse drug reactions. *Genome Med.* **5**, 5 (2013).
31. Wadelius, M. *et al.* Sulfasalazine-induced agranulocytosis is associated with the human leukocyte antigen locus. *Clin. Pharmacol. Ther.* **103**, 843–853 (2018).
32. Png, E. *et al.* A genome-wide association study of hepatitis B vaccine response in an Indonesian population reveals multiple independent risk variants in the HLA region. *Hum. Mol. Genet.* **20**, 3893–3898 (2011).
33. Nishida, N. *et al.* Key HLA-DRB1-DQB1 haplotypes and role of the BTNL2 gene for response to a hepatitis B vaccine. *Hepatology* **68**, 848–858 (2018).
34. Weber, F. *et al.* Single-nucleotide polymorphisms in HLA- and non-HLA genes associated with the development of antibodies to interferon- β therapy in multiple sclerosis patients. *Pharmacogenomics J.* **12**, 238–245 (2012).
35. Perera, M.A., Cavallari, L.H. & Johnson, J.A. Warfarin pharmacogenetics: an illustration of the importance of studies in minority populations. *Clin. Pharmacol. Ther.* **95**, 242–244 (2014).
36. Zhou, Y., Ingelman-Sundberg, M. & Lauschke, V.M. Worldwide distribution of cytochrome P450 alleles: A meta-analysis of population-scale sequencing projects. *Clin. Pharmacol. Ther.* **102**, 688–700 (2017).

37. McInnes, G. *et al.* Pharmacogenetics at Scale: An Analysis of the UK Biobank. *Clin. Pharmacol. Ther.* **109**(6), 1528–1537 (2021).
38. Zhou, Y., Krebs, K., Milani, L. & Lauschke, V.M. Global frequencies of clinically important HLA alleles and their implications for the cost-effectiveness of preemptive pharmacogenetic testing. *Clin. Pharmacol. Ther.* **109**, 160–174 (2021).
39. Yang, J.J. *et al.* Inherited NUDT15 variant is a genetic determinant of mercaptopurine intolerance in children with acute lymphoblastic leukemia. *J. Clin. Oncol.* **33**, 1235–1242 (2015).
40. Relling, M.V. *et al.* Clinical pharmacogenetics implementation consortium guideline for thiopurine dosing based on TPMT and NUDT15 genotypes: 2018 update. *Clin. Pharmacol. Ther.* **105**, 1095–1105 (2019).
41. Daneshjou, R. *et al.* Genetic variant in folate homeostasis is associated with lower warfarin dose in African Americans. *Blood* **124**, 2298–2305 (2014).
42. Johnson, J.A. *et al.* Clinical pharmacogenetics implementation consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin. Pharmacol. Ther.* **102**, 397–404 (2017).
43. De, T. *et al.* Association of genetic variants with warfarin-associated bleeding among patients of African descent. *JAMA* **320**, 1670–1677 (2018).
44. Wen, C.C. *et al.* Genome-wide association study identifies ABCG2 (BCRP) as an allopurinol transporter and a determinant of drug response. *Clin. Pharmacol. Ther.* **97**, 518–525 (2015).
45. Dehghan, A. *et al.* Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet* **372**, 1953–1961 (2008).
46. Rauch, A. *et al.* Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* **138**(1338–45), 1345.e1–7 (2010).
47. Duggal, P. *et al.* Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. *Ann. Intern. Med.* **158**, 235–245 (2013).
48. Vergara, C. *et al.* Multi-ancestry genome-wide association study of spontaneous clearance of hepatitis C virus. *Gastroenterology* **156**, 1496–1507.e7 (2019).
49. Ge, D. *et al.* Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
50. Postmus, I. *et al.* Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins. *Nat. Commun.* **5**, 5068 (2014).
51. Deshmukh, H.A. *et al.* Genome-wide association study of genetic determinants of LDL-c response to atorvastatin therapy: importance of Lp(a). *J. Lipid Res.* **53**, 1000–1011 (2012).
52. Chasman, D.I. *et al.* Genetic determinants of statin-induced low-density lipoprotein cholesterol reduction: the Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial. *Circ. Cardiovasc. Genet.* **5**, 257–264 (2012).
53. Hoffmann, T.J. *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nat. Genet.* **50**, 401–413 (2018).
54. Zhou, K. *et al.* Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. *Nat. Genet.* **48**, 1055–1059 (2016).
55. Wheeler, E. *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Medicine* **14**, e1002383 (2017).
56. Huddart, R. *et al.* Standardized biogeographic grouping system for annotating populations in pharmacogenetic research. *Clin. Pharmacol. Ther.* **105**, 1256–1262 (2019).
57. Li, Q.S., Tian, C., Seabrook, G.R., Drevets, W.C. & Narayan, V.A. Analysis of 23andMe antidepressant efficacy survey data: implication of circadian rhythm and neuroplasticity in bupropion response. *Transl. Psychiatry* **6**, e889 (2016).
58. Yee, S.W. *et al.* Expanding precompetitive multisector collaborations to advance drug development and pharmacogenomics. *Clin. Pharmacol. Ther.* **107**, 96–101 (2020).
59. Verma, S.S. *et al.* Genomewide association study of platelet reactivity and cardiovascular response in patients treated with clopidogrel: A study by the international clopidogrel pharmacogenomics consortium. *Clin. Pharmacol. Ther.* **108**, 1067–1077 (2020).
60. Cirulli, E.T. *et al.* A missense variant in PTPN22 is a risk factor for drug-induced liver injury. *Gastroenterology* **156**, 1707–1716.e2 (2019).
61. Zeggini, E. & Ioannidis, J.P.A. Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10**, 191–201 (2009).
62. Yee, S.W. *et al.* Genomewide Association studies in pharmacogenomics: Meeting report of the NIH pharmacogenomics research network-RIKEN (PGRN-RIKEN) collaboration. *Clin. Pharmacol. Ther.* **100**, 423–426 (2016).
63. Giacomini, K.M. *et al.* Genome-wide association studies of drug response and toxicity: an opportunity for genome medicine. *Nat. Rev. Drug Discov.* **16**, 1 (2017).
64. Schilsky, R.L., McIntyre, O.R., Holland, J.F. & Frei, E. 3rd A concise history of the cancer and leukemia group B. *Clin. Cancer Res.* **12**, 3553s–3555s (2006).
65. Ingle, J.N. *et al.* Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. *J. Clin. Oncol.* **28**, 4674–4682 (2010).
66. Ingle, J.N. *et al.* Genetic polymorphisms in the long noncoding RNA MIR2052HG offer a pharmacogenomic basis for the response of breast cancer patients to aromatase inhibitor therapy. *Cancer Res.* **76**, 7012–7023 (2016).
67. Dudenkov, T.M. *et al.* Anastrozole aromatase inhibitor plasma drug concentration genome-wide association study: Functional epistatic interaction between SLC38A7 and ALPL2. *Clin. Pharmacol. Ther.* **106**, 219–227 (2019).
68. Baldwin, R.M. *et al.* A genome-wide association study identifies novel loci for paclitaxel-induced sensory peripheral neuropathy in CALGB 40101. *Clin. Cancer Res.* **18**, 5099–5109 (2012).
69. Tantisira, K.G. *et al.* Genomewide association between GLCC1 and response to glucocorticoid therapy in asthma. *N. Engl. J. Med.* **365**, 1173–1183 (2011).
70. Himes, B.E. *et al.* Genome-wide association analysis in asthma subjects identifies SPATS2L as a novel bronchodilator response gene. *PLoS Genet.* **8**, e1002824 (2012).
71. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
72. Stark, Z. *et al.* Integrating genomics into healthcare: a global responsibility. *Am. J. Hum. Genet.* **104**, 13–20 (2019).
73. UKB. Data-Field 42040 <<https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=42040>>
74. Hamilton, M. The Hamilton Rating Scale for depression. In *Assessment of Depression* (eds Sartorius, N., Ban, T.A.) 143–152 (Springer, Heidelberg, Berlin 1986).
75. DeBoever, C. *et al.* Assessing digital phenotyping to enhance genetic studies of human diseases. *Am. J. Hum. Genet.* **106**, 611–622 (2020).
76. Sinnott-Armstrong, N. *et al.* Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).
77. Lavertu, A., McInnes, G., Tanigawa, Y., Altman, R.B. & Rivas, M.A. LPA and APOE are associated with statin selection in the UK Biobank. *Cold Spring Harbor. Lab* <https://doi.org/10.1101/2020.08.28.272765>.
78. Martin, A.R. *et al.* Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
79. Jorge, L.F., Eichelbaum, M., Griese, E.U., Inaba, T. & Arias, T.D. Comparative evolutionary pharmacogenetics of CYP2D6 in Ngawbe and Embera Amerindians of Panama and Colombia: role of selection versus drift in world populations. *Pharmacogenetics* **9**, 217–228 (1999).

80. Li, Y.R. & Keating, B.J. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome Med.* **6**, 91 (2014).
81. Atkinson, E.G. *et al.* Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* **53**, 195–204 (2021).
82. Mn, W. *et al.* Use of SNP chips to detect rare pathogenic variants: retrospective, population based diagnostic evaluation. *BMJ* **372**, n214 (2021).
83. Lee, S., Abecasis, G.R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* **95**, 5–23 (2014).
84. Park, J. *et al.* Exome-wide evaluation of rare coding variants using electronic health records identifies new gene-phenotype associations. *Nat. Med.* **27**, 66–72 (2021).
85. Denny, J.C. *et al.* Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1110 (2013).
86. McInnes, G. & Altman, R.B. Drug Response Pharmacogenetics for 200,000 UK Biobank Participants. In *Biocomputing 2021* 184–195 (World Scientific, Hackensack, NJ, 2020).
87. Lanfear, D.E. *et al.* Polygenic score for β -blocker survival benefit in European ancestry patients with reduced ejection fraction heart failure. *Circ. Heart Fail.* **13**, e007012 (2020).
88. He, Y., Hoskins, J.M. & McLeod, H.L. Copy number variants in pharmacogenetic genes. *Trends Mol. Med.* **17**, 244–251 (2011).
89. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
90. Aguirre, M., Rivas, M.A. & Priest, J. Phenome-wide burden of copy-number variation in the UK biobank. *Am. J. Hum. Genet.* **105**, 373–383 (2019).
91. Kolek, M.J. *et al.* A genome-wide association study to identify genomic modulators of rate control therapy in patients with atrial fibrillation. *Am. J. Cardiol.* **114**, 593–600 (2014).
92. Cano-Gamez, E. & Trynka, G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* **11**, 424 (2020).
93. Cairns, J. *et al.* Pharmacogenomics of aromatase inhibitors in postmenopausal breast cancer and additional mechanisms of anastrozole action. *JCI Insight* **5**, e137571 (2020).
94. Suiter, C.C. *et al.* Massively parallel variant characterization identifies NUDT15 alleles associated with thiopurine toxicity. *Proc. Natl. Acad. Sci. USA* **117**, 5394–5401 (2020).
95. Amorosi, C.J. *et al.* Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. *Cold Spring Harbor. Lab* <https://doi.org/10.1101/2021.03.12.435209>.
96. Lauschke, V.M. & Ingelman-Sundberg, M. Emerging strategies to bridge the gap between pharmacogenomic research and its clinical implementation. *NPJ Genom Med* **5**, 9 (2020).
97. McInnes, G. *et al.* Transfer learning enables prediction of CYP2D6 haplotype function. *PLoS Comput. Biol.* **16**, e1008399 (2020).
98. McInnes, G. *et al.* Opportunities and challenges for the computational interpretation of rare variation in clinically important genes. *Am. J. Hum. Genet.* **108**, 535–548 (2021).
99. Lavertu, A. *et al.* Pharmacogenomics and big genomic data: from lab to clinic and back again. *Hum. Mol. Genet.* **27**, R72–R78 (2018).
100. Schaid, D.J., Chen, W. & Larson, N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).
101. Relling, M.V. & Klein, T.E. CPIC: Clinical pharmacogenetics implementation consortium of the pharmacogenomics research network. *Clin. Pharmacol. Ther.* **89**, 464–467 (2011).
102. Kennedy, R.B., Ovsyannikova, I.G., Shane Pankratz, V., Haralambieva, I.H., Vierkant, R.A., & Poland, G.A. Genome-wide analysis of polymorphisms associated with cytokine responses in smallpox vaccine recipients. *Human Genetics.* **131**, 1403–1421 (2012). <http://dx.doi.org/10.1007/s00439-012-1174-2>