Gene Expression **2026** vol. 25(3) | e00021
DOI: 10.14218/GE.2026.00021

Mini Review

PharmacoGWAS for Drug Response: Study Design, Phenotype and Exposure Definition, Bioinformatics Pipelines, Functional Interpretation, and Clinical Translation

Nabil Zaid[1,2*], Dalal Loutfi[2], Lamyaa Benchikhi[1], Banacer Himmi[1], Oussama Badad[3], Hajar El Baroudi[4], Younes Zaid[5], Rajaa Tissir[2] and Hassan Ghazal[2,6]

[1]*Higher Institute of Nursing and Health Sciences Professions (ISPITS), Rabat, Morocco;* [2]*Laboratory of Precision Medicine and One Health, School of Medicine, Mohammed VI University of Sciences and Health, Casablanca, Morocco;* [3]*Southern Illinois University, Carbondale, IL, USA;* [4]*Faculty of Medicine and Pharmacy, Mohammed V University, Rabat, Morocco;* [5]*Department of Biology, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco;* [6]*Department of Sports Sciences, Laboratory of Sports Sciences and Performance Optimization, Royal Institute of Executive Management in Youth and Sport, Sale, Morocco*

Received: June 16, 2026 | **Revised:** July 22, 2026 | **Accepted:** July 27, 2026 | **Published online:** July 29, 2026

Abstract

Inter-individual variability in drug efficacy and toxicity remains a major obstacle to precision therapeutics. Candidate-gene pharmacogenomics and star-allele-based guidelines have established clinically useful examples, but they cannot capture the full spectrum of mechanisms that shape drug response. Pharmacogenomic genome-wide association studies (pharmacoGWAS) extend this framework by enabling discovery beyond known pharmacogenes and can identify human genetic variants associated with efficacy, adverse drug reactions, dose requirements, pharmacokinetics, and pharmacodynamics. This mini-review aims to summarize practical principles for human pharmacoGWAS, with emphasis on study design, phenotype and exposure definition, reproducible bioinformatics pipelines, gene-expression-based functional interpretation, and clinical translation. This review discusses randomized trials, prospective cohorts, biobanks, electronic health records, claims databases, and rare adverse-event designs, highlighting the specific biases that arise because drug response is defined among exposed individuals. It then outlines core analytical steps, including genotype quality control, imputation, ancestry-aware association testing, mixed models, survival and longitudinal analyses, rare-variant aggregation, replication, and meta-analysis. Particular attention is given to expression quantitative trait loci, splicing quantitative trait loci, and protein quantitative trait loci, tissue prioritization informed by the Genotype-Tissue Expression project, transcriptome-wide association studies, and colocalization as tools for prioritizing candidate genes and plausible mechanisms. Finally, we propose a translation framework connecting discovery to clinical validity, guideline development, electronic health record decision support, and equitable implementation across diverse populations. When combined with rigorous epidemiology and functional genomics, pharmacoGWAS may help translate genome-wide signals into safer and more effective prescribing.

Introduction

Drug response is a human phenotype shaped by genetic variation, disease biology, environmental exposures, comorbidities, concomitant medications, and healthcare context. In routine care, variability in response contributes to therapeutic failure, preventable adverse drug reactions (ADRs), and inefficient prescribing. Pharmacogenomics has provided a framework for anticipating some of this variability before treatment is started or while treatment is being optimized. The most mature clinical examples involve variants with large effects in pharmacogenes, including drug-metabolizing enzymes, transporters, human leukocyte antigen (HLA) alleles, and drug targets, for which curated knowledge bases and guideline groups support implementation.[1-3]

This candidate-gene model remains essential, but it is incom-

Keywords: Pharmacogenomics; PharmacoGWAS; Drug response; Gene expression; Functional genomics; Personalized medicine.
***Correspondence to:** Nabil Zaid, Higher Institute of Nursing and Health Sciences Professions (ISPITS), Rabat 10000, Morocco. ORCID: https://orcid.org/0000-0002-9272-6591. Tel: +212-664-777944, E-mail: zaidnabil@hotmail.com
How to cite this article: Zaid N, Loutfi D, Benchikhi L, Himmi B, Badad O, El Baroudi H, *et al*. PharmacoGWAS for Drug Response: Study Design, Phenotype and Exposure Definition, Bioinformatics Pipelines, Functional Interpretation, and Clinical Translation. *Gene Expr* 2026;25(3):e00021. DOI: https://doi.org/10.14218/GE.2026.00021.

plete. Many drug-response phenotypes are polygenic, tissue-specific, time-dependent, and influenced by mechanisms beyond classic absorption, distribution, metabolism, and excretion (ADME) biology. Immune pathways, disease-modifying pathways, gene regulation, drug targets, and compensatory networks can all contribute to efficacy or toxicity. Genome-wide association studies (GWAS) of drug response, or pharmacoGWAS, provide an unbiased strategy to discover these mechanisms in exposed human populations. The increasing availability of biobanks, electronic health records (EHRs), prescription records, national registries, and public genomic resources now makes it possible to study human drug response at scale.[4-6]

Recent studies leveraging these resources have begun to realize this potential: gene-context interaction methods applied to over 340,000 UK Biobank participants have estimated the heritability of response to statins, metformin, warfarin, and methotrexate and identified dozens of modifier genes, while EHR-derived pharmacoGWAS combining UK Biobank and All of Us data have recapitulated established pharmacogenetic loci (*APOE*, *LPA*, *SLCO1B1*) and uncovered a novel rare-variant association (*GIMAP5*) with metformin response, demonstrating that unbiased, biobank-scale discovery may extend beyond classic ADME genes to a broader, polygenic architecture of drug efficacy.[7,8]

This mini-review aims to provide a concise, clinically oriented and gene-expression-aware framework for pharmacoGWAS. We focus on human study designs, phenotype and exposure modeling, bioinformatics and statistical workflows, functional interpretation through molecular quantitative trait loci (QTLs) and transcriptomic resources, and the transition from association signals to clinical implementation. The integrated framework is summarized in Figure 1. As a narrative rather than a systematic review, this article does not follow a pre-specified search protocol, but selectively summarizes methodological and translational literature relevant to human pharmacoGWAS.

Our perspective is deliberately practical. We do not treat pharmacoGWAS as a purely statistical extension of disease GWAS, because drug response requires simultaneous modeling of genetics, treatment exposure, clinical context, and biological mechanism. We also emphasize that a genome-wide signal should not be viewed as clinically useful until it can be connected to a credible gene, pathway, tissue, or prescribing decision. This focus emphasizes how genetic variation influences human gene regulation and, ultimately, medically relevant phenotypes.

Human drug-response phenotypes and study designs

A pharmacoGWAS begins with a clinically meaningful drug-response phenotype. Outcomes may reflect efficacy, toxicity, dose requirements, pharmacokinetics, pharmacodynamics, or treatment persistence. Efficacy endpoints can be quantitative, such as change in HbA1c after metformin initiation; categorical, such as responders versus nonresponders; longitudinal, such as blood pressure trajectories; or time-to-event, such as relapse after treatment initiation. Safety endpoints range from common laboratory abnormalities to rare, severe ADRs, including drug-induced liver injury, severe cutaneous reactions, or drug-associated myopathy.[9,10]

Unlike a conventional disease GWAS, a drug-response GWAS is conditioned on exposure. The analytic population is therefore determined not only by biology but also by prescribing decisions, access to care, adherence, dose changes, comedications, and treatment discontinuation. This creates opportunities for human pharmacogenomic discovery, but also introduces bias. Confounding by indication, channeling bias, survivor bias, immortal time bias, and informative censoring may distort associations if exposure windows and follow-up are poorly specified. A target-trial mindset is useful: investigators should define eligibility, treatment initiation, baseline covariates, treatment strategy, follow-up, endpoint, and analysis plan before selecting genetic models.[11]

Randomized controlled trials remain highly valuable for pharmacoGWAS because treatment allocation, dose schedules, endpoint ascertainment, and follow-up are standardized. Their limitations typically include modest sample sizes, narrow eligibility criteria, and limited power for rare variants or rare ADRs. Prospective cohorts and pharmacokinetic studies provide richer mechanistic phenotyping but require careful control of confounding. Biobank and EHR-linked cohorts offer scale and broad clinical coverage, especially when prescription records, laboratory tests, diagnoses, and outcomes can be linked to genotypes.[5,6] For rare severe ADRs, enriched case-control or case-only designs may be efficient, but ancestry, exposure opportunity, and control definition must be handled with exceptional care.

Phenotype and exposure modeling in real-world human data

Real-world pharmacoGWAS depends on the quality of phenotype and exposure engineering. Prescription records, dispensing data, or claims can be used to define drug exposure, but each source measures a different construct: treatment intent, medication dispensing, or reimbursed therapy. Exposure can be modeled as ever versus never use, new-user status, cumulative dose, average dose intensity, treatment duration, time-varying exposure, or adherence proxy. New-user designs reduce bias from prior treatment, while active-comparator designs can reduce confounding by indication by comparing drugs used for similar clinical conditions. Endpoint algorithms should combine clinical relevance with reproducibility. For efficacy, repeated laboratory or clinical measurements can support longitudinal modeling, allowing baseline values, time since treatment initiation, and treatment changes to be incorporated. For safety, diagnostic codes alone may be insufficient; specificity often improves when codes are combined with laboratory abnormalities, medication discontinuation, hospitalization, or specialist-confirmed events. Validation through chart review or comparison with established phenotyping algorithms increases confidence and helps quantify positive predictive value.

Covariate adjustment should include age, sex, ancestry principal components, genotyping batch, baseline disease severity, renal or hepatic function when relevant, comedications, and clinical factors influencing prescribing. Propensity scores, inverse probability weighting, negative-control analyses, and sensitivity analyses can improve robustness to confounding, even though inherited variants are fixed at conception. Gene-environment correlation can still occur when prescribing differs by ancestry, geography, socioeconomic status, or healthcare access. Table 1 provides a practical checklist for EHR-based pharmacoGWAS.

Time zero is especially important in EHR-based analyses. If baseline is defined after treatment initiation, early failures or early ADRs may be missed, creating immortal time or survivor bias.

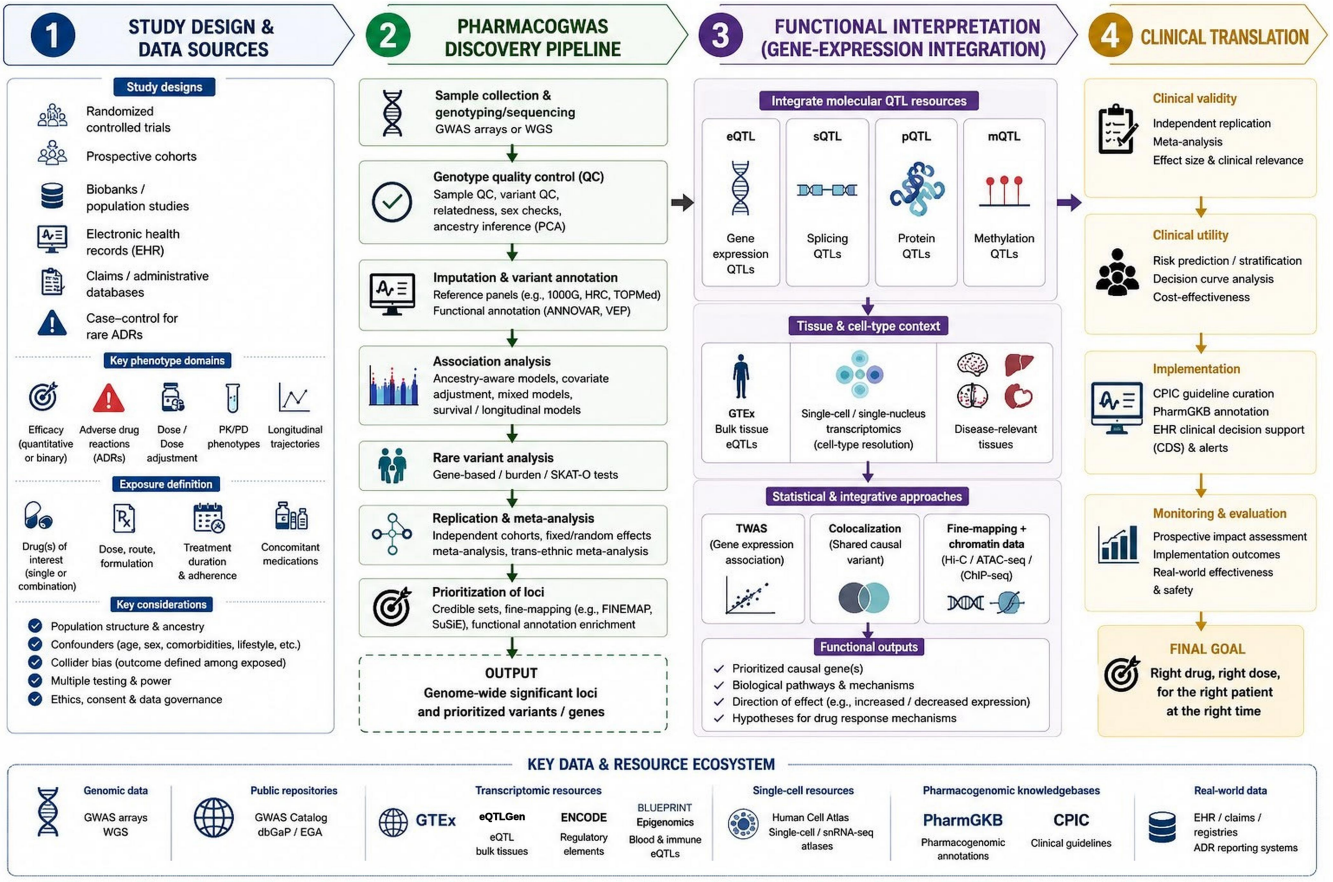

Fig. 1. Overview of the pharmacoGWAS workflow from study design to clinical translation. The figure summarizes the major stages of pharmacogenomic genome-wide association studies (pharmacoGWAS) for investigating human drug response. Phase 1 illustrates study design and data sources, including randomized controlled trials, prospective cohorts, biobanks, electronic health records (EHRs), and administrative databases, together with key considerations for phenotype definition, drug exposure assessment, population structure, confounding, and study governance. Phase 2 presents the pharmacoGWAS discovery pipeline, encompassing sample collection and genotyping or sequencing, quality control (QC), genotype imputation, association analyses, rare variant analyses, replication, meta-analysis, and locus prioritization to identify genome-wide significant variants and candidate genes. Phase 3 highlights the functional interpretation of pharmacoGWAS signals through integration of molecular quantitative trait locus (QTL) resources, including expression quantitative trait loci (eQTLs), splicing QTLs (sQTLs), protein QTLs (pQTLs), and methylation QTLs (mQTLs), together with transcriptomic resources such as the Genotype-Tissue Expression (GTEx) project, single-cell transcriptomics, transcriptome-wide association studies (TWAS), colocalization analyses, and chromatin accessibility data to prioritize candidate genes and plausible biological mechanisms. Phase 4 illustrates the translation of pharmacoGWAS findings into clinical practice through independent replication, functional validation, pharmacogenomic knowledge bases, clinical guidelines, electronic clinical decision support, and post-implementation evaluation to support precision medicine. 1000G, 1000 Genomes Project; ADR, adverse drug reaction; ANNOVAR, ANNOtate VARiation; ATAC-seq, assay for transposase-accessible chromatin using sequencing; CDS, clinical decision support; ChIP-seq, chromatin immunoprecipitation sequencing; CPIC, Clinical Pharmacogenetics Implementation Consortium; dbGaP, database of Genotypes and Phenotypes; EGA, European Genome-phenome Archive; ENCODE, Encyclopedia of DNA Elements; eQTLGen, eQTLGen Consortium; FINEMAP, Fine-mapping of Causal Variants; GWAS, genome-wide association study; Hi-C, high-throughput chromosome conformation capture; HRC, Haplotype Reference Consortium; PCA, principal component analysis; PharmGKB, Pharmacogenomics Knowledgebase; PK/PD, pharmacokinetics/pharmacodynamics; SKAT-O, optimal sequence kernel association test; snRNA-seq, single-nucleus RNA sequencing; SuSiE, Sum of Single Effects; TOPMed, Trans-Omics for Precision Medicine; VEP, Variant Effect Predictor; WGS, whole-genome sequencing.

Conversely, if baseline is too broad, previous treatment, disease progression, or unmeasured exposure may obscure the pharmacogenomic effect. Dose-response analyses may add biological plausibility, but only when dose reflects the intended pharmacological exposure rather than a reaction to early response or toxicity. Therefore, exposure modeling should be developed with clinical experts and, when possible, supported by sensitivity analyses using alternative windows, comparators, and endpoint definitions.

Bioinformatics and statistical pipeline

The computational pipeline must be transparent, reproducible, and adapted to the endpoint. Standard GWAS quality control includes sample call rate, heterozygosity, sex checks, relatedness, ancestry

Table 1. Practical checklist for electronic health record-based pharmacoGWAS

Step	Key considerations
Cohort definition	Define exposed population, eligibility, baseline, comparator group, follow-up, and exclusion criteria
Exposure modeling	Specify initiation, discontinuation, dose, duration, adherence proxies, washout periods, and time-varying exposure
Endpoint algorithm	Use reproducible definitions combining diagnoses, laboratory values, procedures, medication changes, and validation when possible
Bias and confounding control	Address confounding by indication, channeling bias, informative censoring, comedications, and disease severity
Genetic data processing	Apply sample and variant quality control, ancestry inference, relatedness handling, imputation, and allele harmonization
Association model	Match the model to the endpoint: linear, logistic, Cox, longitudinal, mixed-model, interaction, or rare-variant tests
Replication and meta-analysis	Replicate in an independent cohort, harmonize definitions, evaluate heterogeneity, and document effect directions
Functional interpretation	Integrate annotation, fine-mapping, eQTL/sQTL/pQTL evidence, TWAS, colocalization, and pathway analyses
Clinical translation	Evaluate clinical validity, effect size, actionability, guideline readiness, EHR decision support, and equity

This checklist was synthesized from the methodological and translational literature discussed in this mini-review and is not derived from a single dataset. EHR, electronic health record; eQTL, expression quantitative trait locus; pQTL, protein quantitative trait locus; sQTL, splicing quantitative trait locus; TWAS, transcriptome-wide association study.

inference, batch effects, and variant-level filters. Population structure should be modeled using principal components or more advanced ancestry-aware approaches, particularly in multi-ancestry analyses.[12] Appropriate ancestry adjustment is particularly important in multi-ancestry pharmacogenetic GWAS of drug response because it can reduce confounding by population structure and improve detection of population-specific signals that may be missed in analyses restricted to a single ancestry group. For example, a pharmacogenetic study conducted in a multi-ancestry cohort at risk of type 2 diabetes applied this type of adjustment to examine acute response to glucose-lowering treatments, and reported several ancestry-specific variants that had not been reported in previous studies conducted predominantly in European populations, illustrating the value of this approach for discovering relevant variants across diverse populations.[13]

Imputation increases variant coverage and facilitates harmonization across genotyping arrays and cohorts, but requires consistent genome builds, allele alignment, strand checks, reference-panel documentation, and post-imputation quality thresholds.[14-16]

Association models should follow the endpoint structure rather than the convenience of a default GWAS tool. Linear regression is suitable for many quantitative changes, logistic regression for binary outcomes, Cox or flexible survival models for time-to-event endpoints, and mixed models for repeated measurements. Large biobanks often contain related individuals and imbalanced case-control ratios, especially for ADRs; scalable mixed-model tools such as SAIGE (Scalable and Accurate Implementation of Generalized Mixed Model) and REGENIE improve calibration and computational feasibility.[17,18] Interaction models may be needed for gene-by-drug, gene-by-dose, gene-by-ancestry, or gene-by-disease effects, but they require careful interpretation and larger sample sizes.

Rare variants can be important for drug response because loss-of-function or deleterious missense variants in pharmacogenes, transporters, and targets may have large effects. Sequencing or high-quality imputation enables burden tests and variance-component methods using biologically informed masks, such as predicted loss-of-function, damaging missense, or regulatory variants in a tissue-specific enhancer.[19] Replication in an independent cohort is essential. When individual-level replication is not possible, harmonized summary-statistics meta-analysis can improve power, provided that phenotype, exposure, allele coding, genome build, and covariates are carefully aligned.[20]

Transparent reporting is also part of the pipeline. A pharmacoGWAS manuscript should specify the source of exposure data, phenotype algorithms, quality-control thresholds, genome build, imputation reference panel, ancestry definition, relatedness strategy, covariates, multiple-testing threshold, software versions, and replication plan. This information is not merely technical; it determines whether a finding can be reproduced, meta-analyzed, functionally interpreted, or converted into clinical evidence. Where possible, analyses should use version-controlled scripts, containerized environments, and harmonized summary-statistic formats.

Gene-expression-based functional interpretation of pharmacoGWAS signals

The most informative contribution of pharmacoGWAS lies not only in identifying associated loci, but also in explaining how those loci affect human gene regulation and drug-response biology. Many GWAS signals are noncoding and may alter

transcription factor binding, enhancer activity, chromatin accessibility, mRNA splicing, transcript abundance, or protein levels. Functional annotation tools such as Ensembl VEP (Variant Effect Predictor), ANNOVAR (Annotate Variation), FUMA (Functional Mapping and Annotation of GWAS), MAGMA (Multi-marker Analysis of GenoMic Annotation), and LD (Linkage Disequilibrium) score regression can prioritize candidate variants, genes, and pathways, but gene-expression evidence is often needed to move from locus to mechanism.[21-25] Fine-mapping tools such as FINEMAP can further refine candidate causal variants.[26] Population-scale reference resources further support this prioritization step: the Genome Aggregation Database (gnomAD), which aggregates sequencing data from more than 140,000 individuals, provides a continuous, gene-level metric of intolerance to loss-of-function variation that can help distinguish genes under strong selective constraint from those that more readily tolerate disruption, complementing locus-based annotation when prioritizing candidate pharmacogenes.[27]

Expression quantitative trait loci (eQTLs), splicing QTLs (sQTLs), protein QTLs (pQTLs), and other molecular QTLs provide a bridge between inherited variation and molecular function. Resources such as the Genotype-Tissue Expression (GTEx) project characterize genetic regulation of gene expression and splicing across multiple human tissues, enabling investigators to ask whether a drug-response locus influences gene expression in pharmacologically relevant tissues.[28,29] Tissue selection should be guided by drug biology: liver and intestine for metabolism and first-pass exposure, kidney for excretion, blood and immune cells for inflammatory or HLA-mediated ADRs, vascular tissues for cardiovascular drugs, and brain or peripheral nerve tissue for neuropsychiatric or analgesic response.

Colocalization analyses test whether a pharmacoGWAS signal and a molecular QTL signal are likely to share the same causal variant, thereby reducing the risk of attributing a locus to a nearby gene only because of physical proximity.[30] Transcriptome-wide association studies (TWAS), including approaches such as PrediXcan and summary-based TWAS, test whether genetically predicted gene expression is associated with a phenotype.[31,32] In pharmacoGWAS, TWAS can help prioritize genes whose genetically regulated expression is associated with drug efficacy or toxicity. However, TWAS and colocalization should be interpreted jointly: a TWAS association may arise from linkage disequilibrium or model misspecification, while colocalization depends on accurate fine-mapping and appropriate reference panels.

Recent methodological work has addressed this need for joint interpretation by developing a Bayesian framework that combines TWAS and colocalization evidence into a single gene-level probability of causal involvement, rather than treating the two analyses as independent filters. In this approach, colocalization evidence is used to constrain and down-weight TWAS associations that lack support for a shared causal variant, which is intended to reduce spurious TWAS signals driven by linkage disequilibrium between genetically predicted expression and the causal GWAS variant. At the same time, because the method preserves and propagates the uncertainty from both sources of evidence rather than relying on hard significance thresholds, it can retain greater statistical power than colocalization analysis used alone, while still allowing rigorous control of the false discovery rate among implicated genes.[33]

Gene-expression interpretation is particularly relevant for ADME genes, transporters, drug targets, and immune-response genes. A regulatory variant that reduces the expression of a hepatic transporter may alter systemic exposure; a variant influencing the expression of a drug target may change pharmacodynamic response, and an immune-cell eQTL may be associated with susceptibility to hypersensitivity or inflammatory toxicity. Integrating pharmacoGWAS with GTEx, eQTLGen, pQTL resources, chromatin annotations, and drug-gene databases can therefore generate mechanistic hypotheses that are testable in cellular systems, organoids, pharmacokinetic studies, or clinical replication cohorts. This functional layer is also essential for distinguishing discovery signals that are biologically plausible from signals that are statistically significant but clinically uncertain.

Several cautions are important. Molecular QTL effects are context-dependent: an eQTL effect measured in bulk liver or whole blood may not represent the relevant cell state during inflammation, cancer therapy, pregnancy, organ failure, or acute drug exposure. Gene expression may also be altered by the drug itself, while germline eQTL resources usually capture baseline regulation. For that reason, functional interpretation should distinguish inherited regulation from drug-induced transcriptional response. When available, pharmacogenomic cell-line screens, primary-cell perturbation experiments, induced pluripotent stem cell models, organoids, and single-cell atlases can complement population QTL resources by providing cell-type, exposure, and dose context. These approaches are particularly valuable for loci that appear to act through regulatory mechanisms but lack a clear coding candidate.

A practical post-GWAS interpretation workflow should therefore begin with fine-mapping and credible-set definition, continue with variant annotation and tissue-specific regulatory context, then evaluate QTL colocalization and transcriptome-based associations, and finally integrate drug-gene knowledge. Evidence is strongest when several layers point to the same gene and mechanism. For instance, a locus associated with toxicity becomes more convincing if the credible variant lies in a liver enhancer, colocalizes with hepatic expression of a transporter, is supported by a TWAS signal, and is biologically consistent with drug exposure. Conversely, a nearby coding gene should not automatically be assumed causal if the regulatory evidence points to another gene in the same locus.

Two recent studies illustrate how pharmacoGWAS signals can be functionally anchored in gene expression and translated toward clinical application. In a cross-ancestry GWAS of tamoxifen metabolism, Khor *et al.*[34] identified an extended *CYP2D6* locus at chromosome 22q13, in which the majority of associated variants were significantly correlated with hepatic *CYP2D6* expression or enzymatic activity in an independent liver cohort; integrating these regulatory variants with the classical *CYP2D6* activity score substantially improved prediction of active endoxifen concentrations, with potential relevance to individualized tamoxifen dosing in breast cancer. Similarly, a multi-ancestry meta-GWAS of glycemic response to sulfonylureas identified two loci, near *GXYLT1* and within *SLCO1B1*, whose lead variants were identified as eQTLs in blood and liver, respectively; functional

transport assays further supported the role of the *SLCO1B1*-encoded transporter OATP1B1 in hepatic sulfonylurea uptake, and this locus displayed a potentially clinically actionable drug–drug–gene interaction with concomitant statin use. Together, these examples demonstrate how coupling GWAS discovery with tissue-specific eQTL evidence and functional validation can help move pharmacoGWAS findings from statistical association toward mechanistic insight and clinically relevant prediction models.[34,35]

From genome-wide signals to clinical translation

Translation requires more than genome-wide significance. A clinically useful pharmacoGWAS finding should demonstrate reproducibility, biological plausibility, a clinically meaningful effect size, and a feasible action. For many efficacy outcomes, individual variants have smaller effects and may need to be combined with clinical predictors, established pharmacogenes, and polygenic or multi-omic scores.[32,36,37] For some severe ADRs, a single variant with a large effect may warrant consideration of preemptive testing, particularly when alternative drugs are available, the adverse outcome is serious, and sufficient evidence or prescribing guidelines support clinical implementation.[1,2,38]

Guideline development usually requires evidence for clinical validity and clinical utility. The Clinical Pharmacogenetics Implementation Consortium and the Pharmacogenomics Knowledgebase provide models for translating genotype-phenotype evidence into prescribing recommendations and for curating gene-drug relationships.[1,2,38] PharmacoGWAS can expand the evidence base by identifying new loci, refining known mechanisms, or revealing response pathways outside classical pharmacogenes. Nevertheless, actionability depends on whether the genetic result can change prescribing, dosing, monitoring, or patient counseling in a way that improves outcomes. In practice, this translation step remains a major bottleneck: a recent scoping review of pharmacogenomic clinical decision-support tools implemented across 15 institutions found that clinical response rates to genetic alerts ranged widely, from 12% to 73%, while evidence regarding the clinical utility and effectiveness of specific clinical decision support strategies remained limited.[39] These findings indicate that generating actionable genetic associations through pharmacoGWAS is necessary but not sufficient, as the downstream infrastructure needed to translate that evidence into consistent, effective prescribing changes is itself still immature and unevenly implemented.

Implementation also depends on technology. Commercial genome-wide arrays and sequencing panels vary in their pharmacogenomic content, and important ADME variants may be missed, poorly tagged, or difficult to impute.[40-42] If the discovery platform does not capture clinically relevant alleles, downstream reporting and decision support can be incomplete. EHR-integrated clinical decision support requires discrete genotype results, interpretable phenotypes, prescribing triggers, audit trails, clinician education, and governance for updating recommendations.[39] A learning health system can then evaluate whether implementing a pharmacoGWAS-derived rule improves safety, efficacy, and equity in routine care.

Clinical utility should be evaluated in the context of the decision being changed. For an ADR, useful metrics may include absolute risk reduction, number needed to genotype, number needed to treat or harm, availability of alternative drugs, and severity of the event. For efficacy, the key question is whether genotype-guided prescribing improves response beyond standard clinical predictors. For dosing, the value of a variant depends on whether it improves initial dose selection, monitoring intensity, or dose-adjustment algorithms. PharmacoGWAS findings are therefore most likely to translate when the genetic result is available before prescribing and when a clear alternative action exists.

Gaps, equity, and practical recommendations

Several challenges limit pharmacoGWAS. Exposed sample sizes are often small relative to disease GWAS, especially for specific drugs, rare ADRs, or dose-adjusted endpoints. Phenotypes may be misclassified when exposure, adherence, or outcome data are incomplete. Ancestry diversity remains limited in many genomic resources, reducing discovery power and limiting the transferability of predictive models. Drug-response biology is also context-dependent: the same variant may have different effects according to indication, dose, comedication, organ function, or disease stage.

Practical recommendations follow from these limitations. Investigators should start with a clinically meaningful question; define exposure, baseline, follow-up, and endpoint before analysis; prefer new-user or active-comparator designs when possible; validate phenotypes; report ancestry and relatedness handling; select association models according to endpoint structure; prespecify replication; and integrate functional interpretation early rather than treating it as an optional post hoc step. Noncoding and regulatory mechanisms should be prioritized explicitly. A pharmacoGWAS locus becomes more compelling when it can be linked to altered gene expression, splicing, protein abundance, tissue-specific regulation, or drug-target biology.

Equity should be built into study design and translation. Multi-ancestry analyses, ancestry-specific calibration, diverse imputation panels, and local validation are necessary to avoid widening disparities in precision prescribing. This is especially important for populations that remain underrepresented in pharmacogenomic studies and for health systems with limited sequencing infrastructure.

Future pharmacoGWAS will likely be strengthened by integration with longitudinal multi-omics, single-cell regulatory maps, pharmacokinetic modeling, and drug-target genetics. However, increasing complexity should not replace careful clinical thinking. The most useful studies will be those in which the drug, phenotype, exposure window, tissue, gene-expression resource, statistical model, and implementation pathways are aligned from the beginning. Such alignment can help the field move from isolated association signals toward reproducible, interpretable, and equitable precision-prescribing tools.

Limitations

Several limitations of current molecular QTL and transcriptomic resources should be acknowledged when interpreting pharmacoGWAS findings. First, most publicly available eQTL catalogs, including GTEx, are derived from steady-state adult tissues collected opportunistically or postmortem, and therefore capture

baseline regulatory variation rather than the pharmacologically induced transcriptional states relevant to drug response. Second, a substantial fraction of regulatory variants are context-dependent: so-called "response eQTLs" are only detectable following a specific perturbation, such as immune activation or exposure to a therapeutic agent, and are largely invisible in unstimulated tissue.[43] This has been directly demonstrated in human hepatocytes, where joint eQTL mapping across baseline and six drug treatments (rifampin, phenytoin, carbamazepine, dexamethasone, phenobarbital, and omeprazole) increased eQTL detection power by 2.7-fold relative to baseline alone and identified 2,988 response eQTLs, including regulatory variants for the drug-metabolizing gene *CYP3A5* that were undetectable at baseline or in GTEx liver data. Notably, these response eQTLs were preferentially enriched at transcription factor binding sites for Nrf2/GABPA, REST, and HNF4A, suggesting that drug-induced transcription factor recruitment may be an important mechanism contributing to condition-specific regulatory effects that bulk, unstimulated eQTL catalogs cannot capture.[44] A related example is seen with statin exposure, which reveals condition-specific eQTLs at the *GATM* locus, illustrating that baseline-only eQTL data may miss context-dependent regulatory effects.[45] Third, tissue specificity remains difficult to resolve, as many bulk tissue eQTLs are shared across multiple tissues rather than being truly tissue-restricted, and bulk expression data from heterogeneous tissue samples can obscure cell type-specific regulatory effects relevant to a given drug's mechanism of action. In addition, this mini-review is narrative and selective rather than systematic; therefore, it does not provide a comprehensive or quantitatively synthesized assessment of all pharmacoGWAS studies. Taken together, these limitations suggest that baseline, steady-state eQTL resources should be interpreted as a partial and potentially conservative proxy for the drug-induced regulatory landscape, and that dynamic, exposure-specific, and cell type-resolved transcriptomic approaches are needed to capture pharmacogenomically relevant gene regulation.

Conclusions

PharmacoGWAS extends human pharmacogenomics beyond candidate pharmacogenes by enabling unbiased discovery of genetic determinants of efficacy, toxicity, dose requirements, pharmacokinetics, and pharmacodynamics. Its value depends on rigorous human study design, careful exposure and phenotype modeling, reproducible bioinformatics, endpoint-appropriate statistical analysis, and independent replication. The most important translational step is functional interpretation: eQTLs, sQTLs, pQTLs, GTEx-informed tissue prioritization, TWAS, and colocalization can help link noncoding loci to candidate genes and plausible biological mechanisms. When these layers are integrated with clinical evidence, guidelines, EHR decision support, and equity-focused implementation, pharmacoGWAS can help convert genome-wide signals into safer and more effective personalized prescribing.

Acknowledgments

The authors used ChatGPT (GPT-5.6) to assist with language editing, structural refinement, and the generation of Figure 1 based on the authors' original concept and scientific specifications during manuscript preparation. All scientific content, interpretations, references, and final wording were critically reviewed, verified, and approved by the authors, who take full responsibility for the integrity and accuracy of the manuscript.

Funding

This work received no specific funding.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Author contributions

Conceptualization (NZ, HG), methodology (NZ), writing – original draft (NZ), project administration (NZ), investigation (DL, LB, YZ), writing – review and editing (DL, LB, BH, HEB, OB, YZ, RT, HG), validation (BH, HEB, OB), resources (YZ), supervision (RT, HG), visualization (HG). All authors reviewed and approved the final manuscript.

References

[1] Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clin Pharmacol Ther 2011;89(3):464–467. DOI: 10.1038/clpt.2010.279, PMID: 21270786.

[2] Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, *et al*. Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther 2012;92(4):414–417. DOI: 10.1038/clpt.2012.96, PMID: 22992668.

[3] Pirmohamed M. Pharmacogenomics: current status and future perspectives. Nat Rev Genet 2023;24(6):350–362. DOI: 10.1038/s41576-022-00572-8, PMID: 36707729.

[4] Cerezo M, Sollis E, Ji Y, Lewis E, Abid A, Bircan KO, *et al*. The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. Nucleic Acids Res 2025;53(D1):D998–D1005. DOI: 10.1093/nar/gkae1070, PMID: 39530240.

[5] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, *et al*. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562(7726):203–209. DOI: 10.1038/s41586-018-0579-z, PMID: 30305743.

[6] Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, *et al*. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 2023;613(7944):508–518. DOI: 10.1038/s41586-022-05473-8, PMID: 36653562.

[7] Sadowski M, Thompson M, Mefford J, Haldar T, Oni-Orisan A, Border R, *et al*. Characterizing the genetic architecture of drug response using gene-context interaction methods. Cell Genom 2024;4(12):100722. DOI: 10.1016/j.xgen.2024.100722, PMID: 39637863.

[8] Sadler MC, Apostolov A, Cevallos C, Auwerx C, Ribeiro DM, Altman RB, *et al*. Leveraging large-scale biobank EHRs to enhance pharmacogenetics of cardiometabolic disease medications. Nat Commun 2025;16(1):2913. DOI: 10.1038/s41467-025-58152-3, PMID: 40133288.

[9] Daly AK, Donaldson PT, Bhatnagar P, Shen Y, Pe'er I, Floratos A, *et al*. HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. Nat Genet 2009;41(7):816–819. DOI: 10.1038/ng.379, PMID: 19483685.

[10] Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, *et al*. SLCO1B1 variants and statin-induced myopathy--a genomewide study. N Engl J Med 2008;359(8):789–799. DOI: 10.1056/NEJMoa0801936, PMID: 18650507.

[11] Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol 2016;183(8):758–764. DOI: 10.1093/aje/kwv254, PMID: 26994063.

[12] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38(8):904–909. DOI: 10.1038/ng1847, PMID: 16862161.

[13] Li JH, Brenner LN, Kaur V, Figueroa K, Schroeder P, Huerta-Chagoya A, *et al*. Genome-wide association analysis identifies ancestry-specific genetic variation associated with acute response to metformin and glipizide in SUGAR-MGH. Diabetologia 2023;66(7):1260–1272. DOI: 10.1007/s00125-023-05922-7, PMID: 37233759.

[14] Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, *et al*. Next-generation genotype imputation service and methods. Nat Genet 2016;48(10):1284–1287. DOI: 10.1038/ng.3656, PMID: 27571263.

[15] Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, *et al*. A global reference for human genetic variation. Nature 2015;526(7571):68–74. DOI: 10.1038/nature15393, PMID: 26432245.

[16] McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, *et al*. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48(10):1279–1283. DOI: 10.1038/ng.3643, PMID: 27548312.

[17] Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, *et al*. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 2018;50(9):1335–1341. DOI: 10.1038/s41588-018-0184-y, PMID: 30104761.

[18] Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, *et al*. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet 2021;53(7):1097–1103. DOI: 10.1038/s41588-021-00870-7, PMID: 34017140.

[19] Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. Biostatistics 2012;13(4):762–775. DOI: 10.1093/biostatistics/kxs014, PMID: 22699862.

[20] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190–2191. DOI: 10.1093/bioinformatics/btq340, PMID: 20616382.

[21] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, *et al*. The Ensembl Variant Effect Predictor. Genome Biol 2016;17(1):122. DOI: 10.1186/s13059-016-0974-4, PMID: 27268795.

[22] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38(16):e164. DOI: 10.1093/nar/gkq603, PMID: 20601685.

[23] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8(1):1826. DOI: 10.1038/s41467-017-01261-5, PMID: 29184056.

[24] de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11(4):e1004219. DOI: 10.1371/journal.pcbi.1004219, PMID: 25885710.

[25] Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, *et al*. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47(3):291–295. DOI: 10.1038/ng.3211, PMID: 25642630.

[26] Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 2016;32(10):1493–1501. DOI: 10.1093/bioinformatics/btw018, PMID: 26773131.

[27] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581(7809):434–443. DOI: 10.1038/s41586-020-2308-7, PMID: 32461654.

[28] GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 2020;369(6509):1318–1330. DOI: 10.1126/science.aaz1776, PMID: 32913098.

[29] GTEx Consortium; Laboratory; Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, *et al*. Genetic effects on gene expression across human tissues. Nature 2017;550(7675):204–213. DOI: 10.1038/nature24277, PMID: 29022597.

[30] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, *et al*. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10(5):e1004383. DOI: 10.1371/journal.pgen.1004383, PMID: 24830394.

[31] Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, *et al*. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 2015;47(9):1091–1098. DOI: 10.1038/ng.3367, PMID: 26258848.

[32] Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, *et al*. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016;48(3):245–252. DOI: 10.1038/ng.3506, PMID: 26854917.

[33] Okamoto J, Wang L, Yin X, Luca F, Pique-Regi R, Helms A, *et al*. Probabilistic integration of transcriptome-wide association studies and colocalization analysis identifies key molecular pathways of complex traits. Am J Hum Genet 2023;110(1):44–57. DOI: 10.1016/j.ajhg.2022.12.002, PMID: 36608684.

[34] Khor CC, Winter S, Sutiman N, Mürdter TE, Chen S, Lim JSL, *et al*. Cross-Ancestry Genome-Wide Association Study Defines the Extended CYP2D6 Locus as the Principal Genetic Determinant of Endoxifen Plasma Concentrations. Clin Pharmacol Ther 2023;113(3):712–723. DOI: 10.1002/cpt.2846, PMID: 36629403.

[35] Dawed AY, Yee SW, Zhou K, van Leeuwen N, Zhang Y, Siddiqui MK, *et al*. Genome-Wide Meta-analysis Identifies Genetic Variants Associated With Glycemic Response to Sulfonylureas. Diabetes Care 2021;44(12):2673–2682. DOI: 10.2337/dc21-1152, PMID: 34607834.

[36] Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun 2019;10(1):1776. DOI: 10.1038/s41467-019-09718-5, PMID: 30992449.

[37] Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience 2019;8(7):giz082. DOI: 10.1093/gigascience/giz082, PMID: 31307061.

[38] Caudle KE, Klein TE, Hoffman JM, Muller DJ, Whirl-Carrillo M, Gong L, *et al*. Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium

(CPIC) guideline development process. Curr Drug Metab 2014;15(2): 209–217. DOI: 10.2174/1389200215666140130124910, PMID: 24479687.

[39] Smith DM, Wake DT, Dunnenberger HM. Pharmacogenomic Clinical Decision Support: A Scoping Review. Clin Pharmacol Ther 2023; 113(4):803–815. DOI: 10.1002/cpt.2711, PMID: 35838358.

[40] Zaid N, Limami Y, Senhaji N, Errafiy N, Khalki L, Bakri Y, *et al*. Coverage rate of ADME genes from commercial sequencing arrays. Medicine (Baltimore) 2019;98(3):e13975. DOI: 10.1097/MD.0000000000013975, PMID: 30653102.

[41] Zaid N, Benchikhi L, El Abboudi N, Khalki L, Badad O, Limami Y, *et al*. Assessing ADME gene coverage: An observational study on chloroquine therapy for COVID-19. Medicine (Baltimore) 2025; 104(37):e44340. DOI: 10.1097/MD.0000000000044340, PMID: 40958299.

[42] Zaid N, Benchikhi L, Himmi B, Ghazal H. Bioinformatics applications in pharmacogenomics: towards personalized medicine. Gene 2026; 979:149935. DOI: 10.1016/j.gene.2025.149935, PMID: 41352693.

[43] Umans BD, Battle A, Gilad Y. Where Are the Disease-Associated eQTLs? Trends Genet 2021;37(2):109–124. DOI: 10.1016/j.tig.2020.08.009, PMID: 32912663.

[44] Zhong Y, De T, Mishra M, Avitia J, Alarcon C, Perera MA. Leveraging drug perturbation to reveal genetic regulators of hepatic gene expression in African Americans. Am J Hum Genet 2023;110(1): 58–70. DOI: 10.1016/j.ajhg.2022.12.005, PMID: 36608685.

[45] Mangravite LM, Engelhardt BE, Medina MW, Smith JD, Brown CD, Chasman DI, *et al*. A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. Nature 2013;502(7471): 377–380. DOI: 10.1038/nature12508, PMID: 23995691.