Title
Genome-wide association studies in pharmacogenomics: successes and lessons.

Permalink
https://escholarship.org/uc/item/35r0x94z

Journal
Pharmacogenetics and genomics, 23(8)

ISSN
1744-6872

Authors
Motsinger-Reif, Alison A
Jorgenson, Eric
Relling, Mary V
et al.

Publication Date
2013-08-01

DOI
10.1097/FPC.0b013e32833d7b45

Peer reviewed
Genome-wide association studies in pharmacogenomics: successes and lessons
Alison A. Motsinger-Reif, Eric Jorgenson, Mary V. Relling, Deanna L. Kroetz, Richard Weinshilboum, Nancy J. Cox and Dan M. Roden

Objective As genotyping technology has progressed, genome-wide association studies (GWAS) have matured into efficient and effective tools for mapping genes underlying human phenotypes.

Methods Recent studies have shown the utility of the GWAS approach for examining pharmacogenomic traits, including drug metabolism, efficacy, and toxicity.

Results Application of GWAS to pharmacogenomic outcomes presents unique challenges and opportunities.

Conclusion In the current review, we discuss the potential promises and potential caveats of this approach specifically as it relates to pharmacogenomic studies. Concerns with study design, power and sample size, and analysis are reviewed. We further examine the features of successful pharmacogenomic GWAS, and describe consortia efforts that are likely to expand the reach of pharmacogenomic GWAS in the future.

Introduction
Since 2005, genome-wide association studies (GWAS) have matured into a powerful tool to identify single nucleotide polymorphisms (SNPs) that can be reproducibly associated with a variety of human phenotypes. Currently, over 300 papers have reported significant associations of common variants with a range of phenotypes and diseases [1]. These successes have provided numerous insights into the relationship among genetic variants, biological pathways, and human traits, and shown how proper study design and analysis can lead to the success of GWAS. A key lesson from this first generation of GWAS is that no single approach will be appropriate for all phenotypes [2].

The genetics of drug-response outcomes, broadly referred to here as pharmacogenetic/pharmacogenomic outcomes, are a particular category of phenotypes that present unique challenges and opportunities in gene discovery [3]. In this study, we discuss the advantages and limitations of GWAS as applied to pharmacogenomic outcomes. Some of these challenges are variations on general concerns for disease gene identification, whereas others are unique to pharmacogenomic outcomes.

Like studies of disease phenotypes, the success of any pharmacogenomic GWAS will depend on the effect size and allele frequency of genetic variants that influence the trait, the sample size available to detect those variants, the population under study (treatment protocol, dosage, patient features including self-reported race/ethnicity, etc.), and study design (observational study or randomized controlled trial). Unlike most disease phenotypes, pharmacogenomic outcomes often have clear, clinically defined phenotypes and well-understood mechanisms that may underlie variation in drug response, including known systems of transport and metabolism, and sites of drug action. In addition, larger genetic effects may exist for pharmacogenomic traits than for disease phenotypes, providing greater statistical power for genetic association studies.

An important potential limitation for pharmacogenomic GWAS is the sample size. GWAS for traits like height or QT or complex diseases like diabetes need and benefit from large numbers, and currently mega-meta-analyses are identifying and validating associated loci. Such large sample sets are generally not possible for pharmacogenomic outcomes as they usually include by definition both a disease (often with low prevalence) and well-curated drug response phenotype (which further reduces the available study population).

In this study, we discuss key issues for GWAS, including the strengths and limitations of this approach. We then elucidate issues of heightened importance in GWAS of pharmacogenomic traits. We discuss appropriate study designs and analysis strategies, and describe lessons from...
successful pharmacogenomic GWAS. We end with a discussion of ongoing efforts to develop consortia for the purpose of obtaining large sample sizes for drug response outcomes.

Promises
There are clear, well-understood advantages to a genome-wide association approach to phenotype association discovery. GWAS are conventionally intended as an unbiased scan of the genome, interrogating the majority of common genetic variation for disease association. In contrast to a candidate gene approach, whether narrow or broad in scope, GWAS allow the identification of totally novel susceptibility factors that promise to provide us with better biological understanding of phenotypes [4]. There are many candidate mechanisms that drive variability in drug responses: metabolism, transport, targets, target partners, immunologic pathways (e.g., for allergic reactions), etc. that have directed many successful candidate gene studies [5]. However, they cannot identify genes outside the current knowledge of those mechanisms. GWAS allow such novel discovery.

GWAS have distinct advantages as compared with more traditional linkage-based approaches [6]. There are three key general advantages of GWAS approaches for gene identification, each of which are exaggerated for pharmacogenomic outcomes:

1. Case–control cohorts are generally less expensive and easier to collect than extended pedigrees or nuclear families. This is especially true in drug–response studies where it is rare for multiple family members to have well-characterized responses to drug challenges; that is, formal linkage analysis has not been feasible for drug response phenotypes. GWAS do not require the ascertainment of pharmacogenomic interventions in related individuals.

2. Association studies have higher statistical power to detect small-to-modest genetic effects as compared with linkage studies [6]. For pharmacogenomic studies, especially for rare toxicities where sample sizes are limited, this advantage in power may be the difference between success and failure in gene mapping.

3. As linkage disequilibrium (LD) typically stretches over tens of kilobases as opposed to several megabases [6], association signals are more finely localized than linkage signals, which should lead to more rapid identification of causal variants by rapidly narrowing down the regions for follow-up in functional studies – critical for novel mechanistic insights – and, thus, to more rapid translation of findings.

There are additional advantages to GWAS that are more specific to pharmacogenomic outcomes. First, GWAS provide context for understanding the relative importance of genetic contributors to pharmacogenomic traits that may otherwise be unavailable. The genetic component of human phenotypes can be assessed by estimating heritability (the proportion of variation in a trait because of genetic factors) through methods such as variance components analysis, segregation analysis, etc. Each of these methods requires family data, which, as noted above, is usually difficult to collect for pharmacogenomic outcomes [7].

Another specific application of GWAS in pharmacogenomics is the ability to rule out – with prespecified confidence intervals – contributions by unidentified genes to a drug response phenotype. As pharmacogenomic GWAS can directly investigate the role of genetic variation on clinical outcomes, the findings from pharmacogenomic GWAS can be more rapidly translated to clinical practice. As translation to the bedside is one of the goals of pharmacogenomic gene mapping [8], it is important to ensure that any unanticipated important genetic contribution to variability in a drug response is not missed [9]. Of equal importance is the identification of novel mechanisms, both for drug response and/or adverse drug reactions. So, having identified variants in gene X or Y as contributors to a variable drug response, it is key to ensure that there is no other important genetic contributor before mounting a trial. Understanding the influence of genetic variants in drug response can limit unanticipated variability in a drug treatment [9]. The role of GWAS in this process is evident in the evaluation of the genetic component of warfarin dosing [9]. The strong association of variants in VKORC1 and CYP2C9 for stable warfarin dosing was well established [10–12], but before the National Heart, Lung and Blood Institute in the US would mount a large clinical trial it was important to determine whether there were other genetic variants that also had large effects on stable warfarin dosing. GWAS [13,14] have now ruled out large contributions by other loci, thereby allowing clinical trials to proceed [15]. Similarly, a GWAS for clopidogrel effect on ADP-induced platelet aggregation identified only one associated locus, at CYP2C9/19, laying the groundwork for design of clinical trials [16]. As genotyping platforms with increased SNP density become available, the coverage of genetic variation in the human genome will become more complete, providing greater confidence that clinically important genetic effects on pharmacogenomic traits will not be missed. Thus, while many variants in drug metabolism genes have been shown to confer large clinical effects that have often been identified without GWAS (e.g., by well informed candidate gene studies), even GWAS with ‘negative’ results add this crucial additional information [17].

Considerations
Common disease common variant hypothesis
Despite the advantage of GWAS studies discussed above, there are important caveats that must be remembered in their design and application. Although many of these caveats are true of GWAS in general, the impact of these concerns may be different in pharmacogenomic studies than in general trait mapping.
A key assumption in GWAS is what is known as the common disease/common variant hypothesis [18]. The common disease/common variant hypothesis proposes that most of the genetic risk for common, complex diseases is attributable to relatively common [minor allele frequency (MAF) > 0.05] polymorphisms [18]. The alternative to the common disease/common variant hypothesis is that multiple rare variants cause disease at high prevalence in the population through a variety of mechanisms. Such variants can represent genetic heterogeneity of variants in a single gene, or multiple rare variants within genes in the same pathway that have cumulative effects. These two hypotheses have important implications – common variants are thought to impart subtle effects on gene function, often through changes to gene regulation. Rare variants may have larger effects on gene function, such as nonsynonymous variants that alter the amino acid sequence of the resulting protein, and as a result lead to large changes in disease risk or trait values. As a result, it is likely that both common and rare variants will contribute to common phenotypes, but the relative proportions will influence the appropriate methods for discovering associated variants. The GWAS approach is well powered to detect common variants with modest effects. GWAS is less effective in testing rare variation, a problem that is confounded by the DNA microarrays used in these studies, which have been designed to capture common variation. Even ‘next generation’ GWAS that will reliably interroge (directly or indirectly) all variation with MAF greater than 0.005 may be insufficient to identify enough of the contributory variation to allow us to understand biology whether most of that variation has MAF less than 0.005, as the sample sizes required to achieve sufficient statistical power for such effects may be prohibitive. As ‘next generation’ sequencing becomes more accessible, and whole-genome sequencing becomes more affordable, more rare variant analysis will be possible in pharmacogenomics.

Sources of bias
An important concern in GWAS studies for pharmacogenomics is of the potential for bias in the selection of genetic variants [2]. Although large number of variants with low MAF are included in the densest GWAS platforms, GWAS have little power, given sample sizes available, to detect significant associations with low MAF SNPs. In addition, it is widely recognized that genotype quality is not as high for rare variants as it is for more common variants. Consequently, a common approach is to not assess the significance of associations with rare variants (MAF < 0.01). This further compounds the limited statistical power to detect associations with less common genetic variants. Moreover, SNPs included on high-throughput platforms must pass stringent tests for ease of genotyping, which leads regions with gene duplications (and pseudogenes) to be poorly represented on high-throughput genotyping products, and many of these – such as CYPs or the HLA locus – are precisely the genes of greatest interest for pharmacogenetic studies. The human cytochrome P-450 family of genes that encode enzymes active in xenobiotic metabolism have been associated with a large number of pharmacogenetic outcomes [19]. They are known to be highly polymorphic, with a wide range of allele frequencies across populations, and contain complex structural variation with unique haplotypic structure and copy number variations [20]. The coverage of these types of variation is limited on current GWAS genotyping platforms [21].

Study design
Experimental design is a crucial component of any successful GWAS, and pharmacogenomic studies have different advantages and limitations than traditional disease studies. The importance of proper definition and collection of phenotype data has become increasingly appreciated in the context of GWAS [17]. An important advantage in pharmacogenomic studies is that multiple response phenotypes are often collected within the same study, such as efficacy and adverse events, allowing a broader dissection of trait genetics in a single study. However, because all pharmacogenomic outcomes are responses to the environmental exposure of the drug and because these drugs are given in response to a disease condition, there may be complex interactions between disease and drug response relevant in phenotype definition. Precise definitions are essential for both the disease and drug response phenotypes, which are often discrete diagnoses from these complex relationships. For example, in some, but not all cases, rare adverse drug reactions may represent a ‘tail’ of response distributions and where to define that cut-off within the distribution can be a challenge. The SEARCH Collaborative Group showed a successful approach to address this issue by combining patients with both definite and incipient statin-induced myopathy into a single case definition [22]. In other cases, a rare adverse reaction is an unexpected outcome often unrelated to the desired mechanism of action [17].

One efficient use of resources to collect pharmacogenomic phenotypes is to collect samples within the context of clinical trials, which streamlines the collection procedures. The use of clinical trial data for GWAS is not only an efficient use of resources, but has the advantage that similarly treated ‘controls’ for the phenotype of interest are built into the trials. However, because some trials are not designed for GWAS mapping, the study designs used for collection may not be ideal for pharmacogenomic analysis (e.g., multiple drugs used in treatment arms, etc.) [23]. Obviously, this ‘challenge’ is inherent to the treatment of diseases like cancer or end-stage congestive heart failure in which it would be unethical to fail to treat patients with the current standard of care for this illness. If pharmacogenomic efforts are substudies of clinical trials, sample sizes may decrease, which reduces the power...
of the pharmacogenomic component. As meeting recruitment targets is a primary goal in most clinical trials, genomic and pharmacogenomic efforts are often included only as substudies to which patients may or may not consent; as a result, the power and generalizability of genomic studies is compromised. Genetic studies added as an afterthought may be viewed as creating a barrier to recruitment and thus may not be a priority for sponsors. Collecting drug response phenotypes in health care systems with electronic medical records is another method of accruing patients that is now being explored.

Sample size limitations are a challenge in any GWAS, but are amplified in many pharmacogenomic studies. Particularly when studying rare drug reactions or adverse events, it is by definition not feasible to recruit thousands of patients with rare outcomes. This is a particular limitation in pharmacogenomic GWAS, as the replication of association results in independent populations has become the ‘gold standard’ for validation of results [24]. If the collection of a reasonable sample size for a discovery cohort is at the edge of practicality, this makes the collection of a well-powered replication cohort often impossible. Consortia efforts (discussed below) have been motivated by this limitation, to combine samples from across the world to increase power and potentially identify replication cohorts to maximize power and provide validation to associated signals. However, even the establishment of networks of investigators cannot necessarily overcome these limitations, and the field must look for creative approaches of validation/replication possibly involving functional studies or examination of related intermediate phenotypes.

There are unique ‘challenges’ associated with validation/replication for pharmacogenomics. Clinical trials are expensive, and every study is unique as they are designed to represent an advance over earlier studies to answer novel therapeutic questions. Therefore, in pharmacogenomics greater emphasis may have to be placed on functional validation of GWAS ‘signals’ and on biological plausibility. In addition, one must recognize that the larger the sample size, the more likely that features, which confound the genotype/phenotype relationship will be undocumented or uncontrolled; thus diluting the ‘purity’ of the phenotype and potentially reducing the power [25].

Besides sample size, there are other practical limitations in study design for pharmacogenomic studies. As mentioned earlier, family-based designs are generally impractical with drug response outcomes, which mean the field relies heavily on cohort or case–control studies for GWAS [5]. Although the number of cases may be limited by event frequency as discussed above, finding and selecting appropriate controls presents additional challenges. Although GWAS of common diseases have taken advantage of the use of shared controls across studies, this is not often possible in pharmacogenomic studies, as typically controls must also be exposed to the drug of interest (though this may not be necessary in all cases). Other matching criteria must also be considered, such as disease interactions, population admixture, and additional environmental and clinical exposures.

Analysis
As GWAS have become more prevalent, methodologies for the analysis and interpretation of results have coevolved. Many tools have been developed and evaluated in the context of GWAS, and have resulted in the many successes seen to date. However, there are still many challenges in the analysis strategies used for GWAS in general, and particular challenges for pharmacogenomics, as discussed below.

Standard analytical approaches
The majority of earlier GWAS have relied on the use of traditional statistical methodologies for analysis, and several tools have become widely used in the field. Software packages such as PLINK [26] have become very popular in implementing logistic regression (for case–control or cohort studies), linear regression (for quantitative traits), and family-based association tests for GWAS studies.

After various types of corrections for multiple testing (Bonferroni, permutation approaches, etc), results of these analyses are typically prioritized with replication strategies. For single samples, the union of significant results from several analytical approaches (committee-based approaches) or measures of reliability from internal model validation is often used to prioritize robust signals. When more than one sample is available, multistage replication strategies are often employed to discover, prioritize, and validate signals. Finally, when multiple samples are available, meta-analysis is often used to obtain more comprehensive measure of association signals [27]. Challenges in sample collection (discussed above) can limit the use of such multistage replication and meta-analysis strategies in pharmacogenomics. One alternative approach for replication, or at least prioritization, of association signals in pharmacogenetic studies is to use nonclinical GWAS of large collections of human tissue, cell lines, and genetic model organisms [28].

Detecting complex predictive models
Such traditional approaches have been very powerful for identifying strong single-locus associations (low-hanging fruit) for a wide range of phenotypes in both common diseases and pharmacogenomic outcomes (reviewed below), and are typically applied in a way that fits within the ‘unbiased’ intentions of GWAS. Despite the successes of these approaches, their limitations for detecting and prioritizing more complex models have become a hot topic in the literature [29].

As many successful GWAS have been published, the sum of the genetic contributions of associated variants in many common traits is far below the estimated heritability of the
traits. These gaps in explained heritability are potentially clarified by several potential etiologies. Low power to detect low-effect sizes, the presence of rare variants contributing to phenotypes, unmeasured nucleotide or structural variation, complex methylation/epigenetic mechanisms, and gene–gene/gene–environment interactions are all hypothesized to contribute to the unexplained trait variation [29]. In response to these limitations, new analytical approaches are evolving to detect complex genetic risk models discussed below. These limitations are leading to refinement of methods for GWAS analysis, and these may be especially appropriate for pharmacogenomic studies.

**Expert knowledge driven analysis**

Although this ‘unbiased’ intent of GWAS is to detect potential new genetic associations that might not have been considered as candidate genes, there has been a recent appreciation for the fact that these simple analytical approaches ignore the large amount of expert knowledge available for a particular outcome. In response, there has recently been rapid development in the use of network and pathway analysis for analysis of GWAS data [30–33]. Literature searches (automated or hand curated), databases of earlier results, etc. are being exploited to improve the power of GWAS. As it is much known about the mechanism and metabolism of many of the drugs evaluated in pharmacogenomic studies, there is very well-directed guidance for such knowledge-driven analysis. The Pharmacogenomics Knowledge Base [34] is an important resource and data repository that summarizes and curates drug response/gene relationships through gene variant annotation, hand-curated literature review, and important pharmacogenomic genes and pathways. An example of the potential of pathway-based analysis is discussed below.

**Successes in pharmacogenomics**

Arguably the most important demonstrations of the utility and challenges of GWAS in pharmacogenomics are the empirical results of successful studies. A brief description of the outcomes evaluated in pharmacogenomic GWAS and the strongest signals identified is listed in Table 1. Details of each study can be found in the references provided.

The potential and drawbacks of an agnostic, unbiased approach for genetic association studies in pharmacogenetics are illustrated by a GWAS of the activity of a well-known polymorphic drug metabolizing enzyme, thiopurine methyltransferase (TPMT) in lymphoblastoid cell lines from the HapMap project [63]. The goal of the experiment was to assess whether the TPMT polymorphism could be ‘rediscovered’ in this fashion [62]. Although common polymorphisms in TPMT were well tagged, and TPMT polymorphisms were associated with TPMT activity, the GWAS indicated that 96 genes were ranked higher than was TPMT itself. The extent to which these higher ranked genes are false versus true positives is not yet clear, but indicate the difficulty of using GWAS approaches even for putatively monogenic traits.

An example of a GWAS for drug pharmacokinetics is provided by an analysis of methotrexate clearance determined in over 3000 courses of the drug given to 434 children with leukemia [36]. Many candidate gene studies have earlier been conducted to identify genetic variation associated with methotrexate pharmacokinetic variability with limited success. Using GWAS, the SLCO1B1 gene was represented by multiple polymorphisms in several LD blocks, a finding that was replicated in an independent cohort of patients, suggesting that there are multiple mechanisms by which alteration of OATP1B1 (encoded by SLCO1B1) could affect methotrexate pharmacokinetics. Although methotrexate had been shown to be an OATP1B1 substrate, it was a rather weak one [64,65], and so the gene had not risen to the top of candidate gene lists. This finding has implications for both toxicity to methotrexate and to possible drug interactions with widely used OATP1B1 substrates, such as statins.

The utility of pathway-based analysis is shown by Hartford et al. [60], who performed a GWAS examining etoposide-induced leukemia with myeloid/lymphoid or mixed-lineage leukemia. They prioritized variant associations based on expression results, to identify alterations in three biological pathways: adhesion, Wnt signaling, and regulation of actin. Results in an independent validation cohort confirmed the alterations in the adhesion pathway. None of the alterations identified were significantly based on traditional association analysis, showing the potential of more complex modeling to identify pathway-level associations.

Although most of the published studies identified variants at a genome-wide significance level, many of them found strong potential signals that did not stand up to traditional analyses [45,46,66]. These negative results may represent true negative results, but it is highly likely that many of these studies were limited by many of the challenges discussed above (power, coverage, etc).

**Network efforts**

To address many of the limitations discussed above, particularly in regards to limited sample sizes and lack of traditional replication cohorts, researchers are successfully combining resources and establishing worldwide collaborations to support large-scale GWAS. Given the complexities of drug response phenotypes, this approach seems especially appealing in the application of GWAS to pharmacogenomics. By combining cohorts from around the globe, pharmacogenomic studies will have higher power to detect and validate response-determining variants.

The SEARCH Collaborative Group [22] shows the success of such a collaboration. The SEARCH Collaborative Group examined a rare outcome of statin therapy – myopathy,
<table>
<thead>
<tr>
<th>Study</th>
<th>PMID</th>
<th>Trait</th>
<th>Initial sample size</th>
<th>Replication sample size</th>
<th>Region</th>
<th>Reported gene(s)</th>
<th>Strongest SNP-risk allele</th>
<th>P value</th>
<th>Platform (SNPs passing QC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thompson et al.</td>
<td>20031582</td>
<td>Response to statin therapy</td>
<td>1984 individuals</td>
<td>5009 individuals</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>Perlegen (291 998)</td>
</tr>
<tr>
<td>Trevino et al.</td>
<td>19901119</td>
<td>Methotrexate plasma pharmacokinetics and toxicity</td>
<td>434 children</td>
<td>206 children</td>
<td>12p12.2</td>
<td>SLCO1B1</td>
<td>rs11045879-?</td>
<td>1.7 x 10^-10</td>
<td>(500 568)</td>
</tr>
<tr>
<td>Aberg et al.</td>
<td>19975103</td>
<td>Response to antipsychotic therapy (extrapyramidal side effects)</td>
<td>738 schizophrenic individuals</td>
<td>NR</td>
<td>2p12</td>
<td>Intergenic</td>
<td>rs17022444-?</td>
<td>1 x 10^-10 (SAS)</td>
<td>Affymetrix (492 000)</td>
</tr>
<tr>
<td>Garriock et al.</td>
<td>19846067</td>
<td>Response to citalopram treatment</td>
<td>Up to 883 responders, 608 nonresponders</td>
<td>NR</td>
<td>7q26.3</td>
<td>UBE3C</td>
<td>rs69660308-?</td>
<td>4 x 10^-7 (remission)</td>
<td>Affymetrix (430 198)</td>
</tr>
<tr>
<td>Suppiah et al.</td>
<td>19749758</td>
<td>Response to hepatitis C treatment</td>
<td>131 European ancestry responders, 162 European ancestry nonresponders</td>
<td>261 European responders, 294 European nonresponders</td>
<td>19q13.2</td>
<td>IL28A</td>
<td>rs8099917-G</td>
<td>9 x 10^-9 (SAS)</td>
<td>Illumina (311 159)</td>
</tr>
<tr>
<td>Tanaka et al.</td>
<td>19749757</td>
<td>Response to hepatitis C treatment</td>
<td>72 Japanese responders, 82 Japanese nonresponders</td>
<td>122 Japanese responders, 50 Japanese nonresponders</td>
<td>19q13.2</td>
<td>IL28B</td>
<td>rs8099917-G</td>
<td>3 x 10^-32</td>
<td>Affymetrix (621 220)</td>
</tr>
<tr>
<td>Ising et al.</td>
<td>19738353</td>
<td>Response to antidepressant treatment</td>
<td>339 German individuals</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>Illumina (389 251) (pooled)</td>
</tr>
<tr>
<td>Laje et al.</td>
<td>19724244</td>
<td>Response to antidepressant treatment</td>
<td>90 white cases, 90 white controls</td>
<td>30 white cases, 1652 white controls</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>Illumina (100 864)</td>
</tr>
<tr>
<td>Study</td>
<td>Treatment/Condition</td>
<td>Sample Size</td>
<td>SNP</td>
<td>Genotype</td>
<td>p-Value</td>
<td>Platform</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------------------</td>
<td>-------------</td>
<td>----------------------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McClay et al. [43]</td>
<td>Response to antipsychotic treatment</td>
<td>738 cases</td>
<td>4p15.1 Intergenic</td>
<td>rs17390445-?</td>
<td>1 \times 10^{-7}</td>
<td>(Ziprasidone)</td>
<td>Affymetrix &amp; Perlegen (492 900)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shuldiner et al. [16]</td>
<td>Response to clopidogrel therapy</td>
<td>429 Amish individuals</td>
<td>10q24 CYP2C18-CYP2C19-CYP2C9-CYP2C8 cluster</td>
<td>rs12777823-?</td>
<td>10^{-1}</td>
<td>Affymetrix (400 230)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ge et al. [44]</td>
<td>Response to hepatitis C treatment</td>
<td>871 Caucasian, 191 African American, and 75 Hispanic participants</td>
<td>19q13.2 IL28B</td>
<td>rs12979860-C</td>
<td>1 \times 10^{-28}</td>
<td>(combined)</td>
<td>Illumina (565 759)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alkelai et al. [45]</td>
<td>Response to antipsychotic treatment</td>
<td>199 cases, 198 controls</td>
<td>2q24.3 FIGN</td>
<td>rs12476047-C</td>
<td>3 \times 10^{-6}</td>
<td>(combined)</td>
<td>Affymetrix &amp; Perlegen (495 172)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comabella et al. [46]</td>
<td>Response to interferon beta therapy</td>
<td>53 responders, 53 nonresponders</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>Illumina (428 867) (pooled)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teichert et al. [47]</td>
<td>Aacenocoumarol maintenance dosage</td>
<td>1451 Caucasian patients</td>
<td>16 STX4A MYST1 BCKDK RNF40 CTF1 VKORC1 KIA0339 NNN175901 IGAM ITGAL ITGAX GZNF689 PYCARD FUS FBX19 BCKDK FLJ23426 FLJ23436 ITGAX RNDF RNDF SCRAP CYP2C9 GZNF689 PYCARD FUS CYP2C18 CYP2C19 CYP2C9 HOPX, HLA-B</td>
<td>rs10871454-?</td>
<td>2 \times 10^{-23}</td>
<td></td>
<td>Illumina (550 000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daly et al. [48]</td>
<td>Drug-induced liver injury (fluocloxacin)</td>
<td>8 cases, 282 controls</td>
<td>6p21.33 ST6GAL1</td>
<td>rs10937275-?</td>
<td>1 \times 10^{-8}</td>
<td>(B*5701 positive)</td>
<td>Illumina (866 399)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pharmacogenomics GWAS Motsinger-Reif et al. 389
<table>
<thead>
<tr>
<th>Study</th>
<th>PMID</th>
<th>Trait</th>
<th>Initial sample size</th>
<th>Replication sample size</th>
<th>Region</th>
<th>Reported gene(s)</th>
<th>Strongest SNP-risk allele</th>
<th>P value</th>
<th>Platform (SNPs passing QC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perlis et al.</td>
<td>19448189</td>
<td>Response to lithium treatment in bipolar disorder</td>
<td>458 lithium-treated patients, 719 nonlithium treated patients</td>
<td>359 patients</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>Affymetrix (~ 1.4 million)</td>
</tr>
<tr>
<td>Takeucki et al.</td>
<td>19300499</td>
<td>Warfarin maintenance dose</td>
<td>1053 individuals</td>
<td>588 individuals</td>
<td></td>
<td></td>
<td>10q23.33</td>
<td>3 \times 10^{-181}</td>
<td>Illumina (325 997)</td>
</tr>
<tr>
<td>Yang et al.</td>
<td>19176441</td>
<td>Response to treatment for acute lymphoblastic leukemia</td>
<td>487 children</td>
<td></td>
<td>10p1.2</td>
<td></td>
<td>2q33, 10q23.33, 11q21</td>
<td>9 \times 10^{-7}</td>
<td>Affymetrix (476 976)</td>
</tr>
<tr>
<td>French et al.</td>
<td>19066393</td>
<td>Methotrexate polyglutamate intracellular accumulation</td>
<td>248 patients, 176 HapMap cell lines</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>SA16</td>
<td>NA</td>
<td>Affymetrix (447 287)</td>
</tr>
<tr>
<td>Mick et al.</td>
<td>18821564</td>
<td>Methylphenidate efficacy in pediatric attention deficit/hyperactivity disorder treatment</td>
<td>187 children with attention-deficit/hyperactivity disorder</td>
<td>NR</td>
<td>NS</td>
<td>NS</td>
<td>NA</td>
<td>NA</td>
<td>Affymetrix (319 722)</td>
</tr>
<tr>
<td>Link et al.</td>
<td>18650507</td>
<td>Statin-related muscle toxicity</td>
<td>85 cases, 90 controls</td>
<td>19856 individuals</td>
<td>12p1.2</td>
<td>SLCO1B1</td>
<td>rs4149056-C</td>
<td>2 \times 10^{-9}</td>
<td>Illumina (316 184)</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>18615156</td>
<td>Response to TNF antagonist treatment</td>
<td>89 cases</td>
<td></td>
<td>20q12</td>
<td>IAFB</td>
<td>rs6029045-T</td>
<td>2 \times 10^{-7}</td>
<td>Illumina (283 348)</td>
</tr>
<tr>
<td>Sarasquete et al.</td>
<td>18594024</td>
<td>Development of jaw osteonecrosis after bisphosphonate in myeloma</td>
<td>22 cases and 65 matched controls</td>
<td>NR</td>
<td>10q23.33</td>
<td></td>
<td>10q23.33</td>
<td>4.231 \times 10^{-6}</td>
<td>Affymetrix (500 568)</td>
</tr>
<tr>
<td>Turner et al.</td>
<td>18591461</td>
<td>Response to diuretic therapy</td>
<td>194 Blacks, 195 Whites</td>
<td>NR</td>
<td>12q15</td>
<td>LYZ, YEATS5, FRS2</td>
<td>rs10718454-A</td>
<td>6 \times 10^{-6}</td>
<td>Affymetrix (up to 102 334)</td>
</tr>
<tr>
<td>Cooper et al.</td>
<td>18535201</td>
<td>Warfarin maintenance dose</td>
<td>181 individuals</td>
<td>374 individuals</td>
<td>16p11.2</td>
<td></td>
<td>10p23.33</td>
<td>3 \times 10^{-6}</td>
<td>Illumina (538 628)</td>
</tr>
</tbody>
</table>

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
<table>
<thead>
<tr>
<th>Name et al.</th>
<th>PMID</th>
<th>Sample Description</th>
<th>Sample Size</th>
<th>Genes/Regions</th>
<th>SNPs/Markers</th>
<th>p-value</th>
<th>Genotyping Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volpi et al. [56]</td>
<td>18521091</td>
<td>Response to iloperidone treatment (QT prolongation)</td>
<td>183 individuals</td>
<td>10q23.1, NRG3, rs4933824-T</td>
<td>2 x 10^{-6}</td>
<td>Affymetrix (339 272)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14q12, NUBPL, rs7142881-A</td>
<td>2 x 10^{-6}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15q26.1, SLCO3A1, rs3924426-T</td>
<td>2 x 10^{-6}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18q12.2, BRUNOL4, rs4799915-T</td>
<td>3 x 10^{-6}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2q31.3, CERRL, rs999648-T</td>
<td>3 x 10^{-6}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4q32.3, PALD, rs17054392-C</td>
<td>3 x 10^{-6}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lavedan et al. [57]</td>
<td>18521090</td>
<td>Response to iloperidone treatment (PANSS-T score)</td>
<td>106 individuals</td>
<td>4q25.3, SLC6A11, 5q34, GABRB2, 3q26.3, GABRG3</td>
<td>0.00043</td>
<td>Affymetrix (334 563)</td>
<td></td>
</tr>
<tr>
<td>Inada et al. [58]</td>
<td>18334916</td>
<td>Neuroleptic-induced treatment-resistant tardive dyskinesia</td>
<td>50 Japanese schizophrenia patients with treatment-resistant TD and 50 Japanese schizophrenia patients without TD</td>
<td>36 patients with TD and 136 patients without TD</td>
<td>0.00007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3p25.3, SLC6A11, 5q34, GABRB2, 3q26.3, GABRG3</td>
<td>0.00007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Byun et al. [59]</td>
<td>18195134</td>
<td>Response to interferon beta therapy</td>
<td>206 multiple sclerosis cases</td>
<td>10q23.1, HLA-DRB1, DRB1*07</td>
<td>9 x 10^{-6}</td>
<td>Affymetrix (~100 000) (pooled)</td>
<td></td>
</tr>
<tr>
<td>Hartford et al. [60]</td>
<td>17673902</td>
<td>Etoposide-induced secondary leukemia</td>
<td>3 secondary leukemia/myelodysplasia cases and germline DNA from 13 matched and 156 unmatched controls</td>
<td>NR, NR, Genes in adhesion, Wnt signaling and actin regulation [37]</td>
<td>NR, NR, Genes in adhesion, Wnt signaling and actin regulation [37]</td>
<td>Perlegen (~266 722)</td>
<td></td>
</tr>
<tr>
<td>Kindmark et al. [61]</td>
<td>17505501</td>
<td>Ximelagatran-related liver toxicity</td>
<td>74 cases, 130 controls</td>
<td>10 cases, 16 controls, 6p21.3, HLA-DRB1, DRB1*07</td>
<td>9 x 10^{-6}</td>
<td>Perlegen (~266 722)</td>
<td></td>
</tr>
</tbody>
</table>

Columns list the first author of the study, the PubMedID (PMID), the sample size of the initial sample, the replication sample (if applicable), details of the regions and variants that were statistically significant (if applicable), and the genotyping platform (and number of markers evaluated) used. Studies are arranged in reverse chronological order, according to PubMed ID assignment.

GWAS, genome-wide association study; NA, not applicable; NR, not replicated; NS, not significant; PANSS-T, positive and negative syndrome scale total; QC, quality control; SNP, single nucleotide polymorphism; TD, tardive dyskinesia.
defined as markedly elevated creatinine kinase. In its most extreme form, this can result in the potentially fatal adverse effect of rhabdomyolysis, but these cases are exceedingly rare. The SEARCH Collaborative Group also found that myopathy was rare (approximately 0.1%) with low-dose simvastatin, so they focused their efforts on 98 cases identified in 6031 patients receiving high doses (80 mg/day) of the drug. A GWAS that studied 85 of these cases and 90 controls identified rs4363657, in perfect LD with a known functional nonsynonymous SNP in *SLCO1B1* at genome-wide significance. The 5-year incidence of myopathy was 18% in individuals homozygous for the risk allele (2.1% of the study group), 3% in heterozygotes, and 0.6% in those with no risk allele. The result was replicated in a separate cohort of patients receiving a lower dose of 40 mg/day (relative risk 2.6 per C allele).

The success of this study illustrates several important points in the study design of pharmacogenomic GWAS. First, large collaborative samples can provide a valuable resource for collecting a critical mass of patients with a rare phenotype. Second, rare phenotypes are sampled from the extreme tail of drug response distributions. As a result, genetic variants that influence these traits may have larger genetic effect sizes, and therefore be detectable with small sample sizes, than more common outcomes. Third, similar outcomes can sometimes be combined into a single case group. Here, in the initial association phase, definite and incipient myopathy patients were considered together. Fourth, replication of an association should take place in a similar population. In this study, the replication cohort was treated with a lower dose, 40 mg of simvastatin daily as compared with 80 mg in the initial group. We note that selecting cases from lower dose regimen for a follow-up study may be preferable to the converse (i.e. higher doses in the follow-up cohort), as those cases have a more extreme phenotype (by developing toxicity at a lower dose). This can limit the dilution of the association signal in the confirmatory study.

Several additional pharmacogenomics consortia have been established to evaluate a number of drug response outcomes, including the International Severe Irritotoc Neutropenia Consortium (http://www.pharmgkb.org/views/project.jsp?pld = 69), and the International Tamoxifen Pharmacogenomics Consortium (http://www.pharmgkb.org/views/project.jsp?pld = 63). These groups have pooled data from around the world to investigate genetic predictors of drug response with high power and comparison across global populations. Although the initial study of these consortia has typically been focused on candidate-known genetic effects, they are moving towards GWAS. For example, the International Warfarin Pharmacogenetics Consortia–GWAS consortium) to identify and confirm earlier findings, and potentially discover novel variants that explain potential trait variation across multiple populations. Such collaborations are extremely important for rare events, such as adverse events. The International Serious Adverse Events Consortium (www.saconsortium.org) represents one important effort in pharmacogenomics for adverse events, pulling together commercial, academic, and industry partners to collect data for well-powered GWAS.

These combined datasets represent exciting resources for pharmacogenomics GWAS, but are not without challenges. Concerns with consistent data collection, storage, data-ownership issues, etc. can be concerns in these collaborative efforts.

**Conclusion**

GWAS have proven to be an exciting tool for gene mapping in common human traits, and are showing their potential in pharmacogenomic outcomes as well. As pharmacogenomic GWAS mature, there is an increased appreciation for issues that are specifically related to these unique phenotypes. Practical considerations, related to study design and available sample sizes highlight the need for creative methods of replication, beyond the traditional replication cohorts that are used for common disease genetics, and the necessity of combining samples across consortia. The complex physiology of drug response outcomes highlights the need for analytical methods that incorporate this complexity, using the wealth of information available about drug mechanisms and pathways.

**Acknowledgements**

The authors are grateful to the Pharmacogenetics Research Network (PGRN) Publications Committee for helpful advice. Some of the study described in this review was funded by the NIH/NIGMS Pharmacogenetics Research Network grants U01 GM61393, U01 HL65962, U01 HG04603, U01 GM61388, U01 GM63340, U01 GM61390, and Database (U01GM61374 http://www.pharmgkb.org). This study was additionally supported by CA21765, the American Lebanese Syrian Associated Charities (AL-SAC), and by CureSearch. Dr Relling receives a portion of the income St. Jude receives from licensing patent rights related to TPMT polymorphisms. This publication (or project) was partially supported by NIH/NCRR/OD UCSF-CTSI Grant Number KL2 RR024130. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

**Conflicts of interest**

There are no conflicts of interest.

**References**


