Published in final edited form as:

JAMA. 2011 August 10; 306(6): 652–653. doi:10.1001/jama.2011.1125.

Genomic Technology Applied to Pharmacological Traits

Dr. Jun J. Yang, PhD and

Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, Tennessee

Dr. Robert M. Plenge, MD, PhD

Division of Rheumatology, Immunology and Allergy, and Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts

Advances in genomic technologies have enabled a deeper understanding of a variety of phenotypes relevant to clinical medicine. The ability to test hundreds of thousands of inherited genetic markers for association with medical traits (genome-wide association studies, or GWAS) has led to exciting new discoveries. While GWAS of diseases focuses on disease onset and rate of progression, the term pharmacogenomics refers to the application of GWAS and related genomic technologies to understanding variability in treatment outcome of a disease (a drug's efficacy or toxicity).

GWAS of drug efficacy or toxicity

The human genome is riddled with genetic variation. Recent report from the 1,000 Genome Project described >16 million DNA variants that contribute to differences between human individuals¹: sequence substitutions (single nucleotide polymorphisms [SNPs], ~15 million), short insertion/deletions (~1 million, including repeats), and structural variations (~20,000, copy number changes), representing the majority of common polymorphisms (allele frequency >1%). When all rare variants are considered, the number of genetic variation is even larger. If a DNA variant is present at a differential frequency in one group of patients (e.g., those who respond to a drug) compared to another group of patients (e.g., those who do not respond), then a variant is said to be “associated” with clinical outcome. Such genetic associations can be used to provide insight into important biological pathways or to help guide clinical care in a more personalized manner.

A fundamental challenge in human genetics has been to design experiments that comprehensively test human genetic variation for influence on clinical traits, including pharmacological traits. Until recently, it was technically feasible to test only a small number of SNPs within a small number of the ~20,000 genes in the human genome. Although limited in scope, candidate gene studies led to the discovery of DNA variants that influence warfarin dose requirement, clopidogrel response and toxicity to thiopurines, among others².

With a better understanding of patterns of genetic diversity and with improved technical capacity to test (or “genotype”) hundreds of thousands of SNPs across the genome in a high-throughput fashion, GWAS became feasible. Early successes came from GWAS of disease risk. Applied to pharmacological traits, GWAS have also led to important discoveries with potential clinical application^{3–6}. For example, a GWAS of patients with chronic hepatitis C infection identified common SNPs in the *inosine triphosphatase (ITPA)* gene that protect against ribavirin-induced hemolytic anemia⁶. A GWAS for glycemic response to metformin

Correspondence: please address to either J.J. Yang (Jun.Yang@stjude.org) or R.M. Plenge (rplenge@partners.org).

The authors have no conflict of interest to declare.

identified a SNP associated with treatment success at a locus containing *ATM*, the ataxia telangiectasia mutated gene³. Both GWAS revealed unexpected biological pathways, highlighting the importance of an unbiased approach over a candidate gene approach, as our knowledge of a drug's mechanisms of action may be limited, even for widely used medications. Pharmacogenomic studies of warfarin response, on the other hand, provide good examples of how GWAS complement and extend candidate gene studies⁷: it confirmed previously described variants of *CYP2C9* and *VKORC1* but also discovered contribution of other metabolizing enzymes (*CYP4F2*).

Interpretation of GWAS results

GWAS studies are particularly susceptible to false positive findings due to the large number of genetic variants tested (i.e. multiple testing)⁸. Therefore, it is critically important to impose rigorous statistical criteria to ascertain true hits, including (but not limited to) very stringent P-value cutoffs (e.g., $P < 5 \times 10^{-8}$ for genome-wide significance) and replication/validation in independent cohorts. However, P-values should be considered in the context of statistical power of the study (effect size and sample size). Thus, failure to reach the genome-wide level of statistical significance does not automatically rule-out any contribution of the SNP to the phenotype. Finally, biological experimentations that functionally characterize genetic variations (e.g. in cell lines and animal models) are critical for understanding the mechanisms underlying the observed genotype-phenotype associations.

The majority of GWAS platforms currently available include SNPs based on their representation of the entire genome instead of functionality of the variant. As a result, associated SNPs (or “hits”) from GWAS are usually not the causal variant, but rather markers that are highly correlated with the variants causal for the drug response phenotypes (e.g. through linkage disequilibrium, i.e. tight correlation between genotypes at neighboring loci)⁶. Nevertheless, GWAS signals highlight genomic regions of importance and enable focused sequencing and in-depth genotyping to identify the actual gene and its causal variants.

Additionally, rare DNA variants, which are largely responsible for the genetic basis of highly penetrant Mendelian disorders, are poorly represented on contemporary GWAS arrays. New GWAS arrays offer the potential to test low-frequency DNA variants (e.g., frequency 0.1–5% in the general population). However, if a gene harbors an excess number of rare mutations that are private to one group of patients (e.g., those with toxicity) compared to another group of patients (e.g., those without toxicity), then the gene can only be tested for additional rare variants by direct re-sequencing. Until recently, it was too expensive to sequence a large number of genes in the human genome in large patient collections⁹. As the cost of sequencing continues to drop, it is expected that “next generation” sequencing will become a powerful method for pharmacogenetic discovery efforts⁹. In fact, early successes have already started to emerge from using whole-genome (or exome/transcriptome) sequencing to discover genetic underpinning of diseases (e.g. Mendelian diseases as well as cancers)⁹.

Even with well-established pharmacogenetic relationships, genetic variants usually explain a minority of the observed inter-patient variability. Some of the remaining inter-patient variability may be due to common SNPs with a small effect on phenotype (those to be identified only with extremely large sample sizes) and/or rare variants with large effect (those to be found only via sequencing), both of which are usually not well interrogated in conventional GWAS. Other sources of variability might not be coded within the portion of DNA inherited from parents (germline), but rather reflect changes in DNA methylation or

histone modifications (together termed “epigenetic” modifications), somatic mutations that occur during a person’s lifetime (e.g., in tumor cells), and environmental factors. Pharmacogenomics of anticancer agents are complex because genetic variations in both host and tumor genomes are involved in determining anticancer effects of medications. An important example is gefitinib, whose efficacy is related to activating mutations in its molecular target *EGFR* in non-small-cell lung tumor cells¹⁰.

Future challenges

As noted above, a number of drug-related GWAS have clearly shown the feasibility and potential clinical value of pharmacogenomics to understanding variable drug actions. As the field moves beyond the “low-hanging fruit” of common variants with moderate to large effects, future pharmacogenomic studies are likely to require large sample sizes to detect variants with smaller effect sizes or lower frequencies. However, large sample sizes can be difficult to procure for pharmacological traits. One solution to this problem is to form consortia of investigators to pool clinical data, biospecimens, and when possible, GWAS data, such as those set up under the framework of National Institutes of Health Pharmacogenomics Research Network (PGRN, www.pgrn.org). Another approach is to think creatively about alternative methods of collecting phenotypic data. Recent efforts to use electronic medical records and internet-based registries offer hope for a cost-effective strategy to collect sufficient clinical data linked with biospecimens to perform appropriately powered genetic studies¹¹.

Advances in genomic technology have driven approaches to understand variability in pharmacological traits. In a few short years the field has moved from candidate gene association studies to GWAS, and the field is rapidly moving towards next-generation sequencing to test all classes and frequencies of DNA variants, and to test epigenetic changes in the human genome. The initial success portends that future genomic efforts will continue to shed important light on variability in pharmacological traits.

Acknowledgments

JJY is supported by grants from the NIH (R21-CA158568, RC4-CA156449, and U01-GM92666), ALSF, ASCPT, and ALSAC. RMP is supported by grants from the NIH (R01-AR057108, R01-AR056768, U01-GM092691, R01-AR059648), and Burroughs Wellcome Fund.

References

1. A map of human genome variation from population-scale sequencing. *Nature*. Oct 28; 2010 467(7319):1061–1073. [PubMed: 20981092]
2. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *N Engl J Med*. Mar 24; 2011 364(12):1144–1153. [PubMed: 21428770]
3. Zhou K, Bellenguez C, Spencer CC, et al. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet*. Feb; 2011 43(2):117–120. [PubMed: 21186350]
4. McCormack M, Alfievic A, Bourgeois S, et al. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. *N Engl J Med*. Mar 24; 2011 364(12):1134–1143. [PubMed: 21428769]
5. Shuldiner AR, O’Connell JR, Bliden KP, et al. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA*. Aug 26; 2009 302(8): 849–857. [PubMed: 19706858]
6. Fellay J, Thompson AJ, Ge D, et al. ITPA gene variants protect against anaemia in patients treated for chronic hepatitis C. *Nature*. Mar 18; 2010 464(7287):405–408. [PubMed: 20173735]

7. Takeuchi F, McGinnis R, Bourgeois S, et al. A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet.* Mar.2009 5(3):e1000433. [PubMed: 19300499]
8. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA.* Mar 19; 2008 299(11):1335–1344. [PubMed: 18349094]
9. Mardis ER. A decade's perspective on DNA sequencing technology. *Nature.* Feb 10; 2011 470(7333):198–203. [PubMed: 21307932]
10. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med.* May 20; 2004 350(21):2129–2139. [PubMed: 15118073]
11. Wilke RA, Xu H, Denny JC, et al. The emerging role of electronic medical records in pharmacogenomics. *Clin Pharmacol Ther.* Mar; 2011 89(3):379–386. [PubMed: 21248726]