Published in final edited form as:
Clin Pharmacol Ther. 2013 August ; 94(2): 211–213. doi:10.1038/clpt.2013.111.

Challenges in the Pharmacogenomic Annotation of Whole Genomes

RB Altman[1,2], **M Whirl-Carrillo**[1], and **TE Klein**[1]

[1]Department of Genetics, Stanford University, Stanford, California, USA

[2]Department of Bioengineering, Stanford University, Stanford, California, USA

Abstract

Next-generation sequencing technologies have enabled the sequencing of an entire genome for less than $4,000. With access to all the genetic variation in an individual, our task will be to use this information to assess the genetic influences on drug efficacy, toxicity, and dosage.

Next-generation sequencing technologies enable the sequencing of an entire genome for a few thousand dollars, and costs are expected to decrease over the next few years. Therefore, we can expect that an increasing number of patients will have genomic information to inform drug-prescribing decisions. The overall quality of genome sequencing is rapidly improving, but there are still regions of the genome that are difficult accurately to sequence. Unfortunately, some of these regions are very important, and so current efforts are focused on improving the quality of data from these regions. At the same time, interpretation of genome data is difficult and labor intensive. Our task is to use the genome to assess its influences on drug efficacy, toxicity, and dosage. Since 2000, The Pharmacogenomics Knowledgebase (PharmGKB, http://www.pharmgkb.org/) has captured and cataloged published knowledge of human genetic variation and how it impacts drug-response phenotypes.[1] PharmGKB curators manually curate the peer-reviewed literature that establishes associations between variants and their impacts. Curators also create drug-centered pathways of pharmacodynamic and pharmacokinetic action. Today, the PharmGKB has information about nearly 900 genes and 1,200 variations influencing 900 drug responses. There are 95 PharmGKB-curated pathways and 48 summaries of “very important pharmacogenes.” These basic research findings have been synthesized into more than 600 “clinical annotations,” and for 76 gene–drug pairs, there are published dosing guidelines from the Royal Dutch Association for the Advancement of Pharmacy–Pharmacogenetics Working Group[2] and the Clinical Pharmacogenetics Implementation Consortium.[3] These annotations are labeled to indicate the strength of evidence supporting their clinical implementation.[1]

Correspondence: RB Altman (russ.altman@stanford.edu).

CONFLICT OF INTEREST

R.B.A. is a founder of and consultant to Personalis, which focuses on clinical genomics using next-generation sequencing. The other authors declared no conflict of interest.

The PharmGKB team has participated in three efforts to clinically annotate complete human genome sequences. For each, the goal was to assess the potential impact of pharmacogenomics on clinical practice. One paper examined the genome of a 40-year-old asymptomatic man.[4] His genome showed that he was a *CYP2C19*2* carrier, potentially affecting his response to many drugs, including clopidogrel. He was also warfarin insensitive and had an average risk of myopathy in response to statins. The exercise produced clinically relevant information (with levels of evidence varying from low to high) for nearly 100 drugs. Another paper examined the anticipated drug responses for a family of four, focusing on the inheritance pattern of drug response.[5] Finally, a third paper showed the drug response of an individual with simultaneous measurements of the transcriptome, metabolome, and other "omes" over time.[6]

These exercises were useful in defining our current ability to provide clinically useful information, as well as in outlining some outstanding challenges. They combine with recent developments in sequencing and the work of others to suggest a research agenda for pharmacogenomics in the context of whole-genome sequencing:

1. *Accuracy matters and is variable across the genome*. Whole-genome sequencing technologies are changing rapidly, and accuracy varies across the genome. For example, the coding region of many genes is rich in GC bases and is less well interrogated by some platforms. In addition, insertions/deletions are difficult to call accurately with some next-generation sequencing technologies. Copy-number variations are also important for "VIP genes" (such as the cytochrome P450 system) and are difficult to interrogate. Pseudogenes within large gene families also complicate our ability to reliably assess variation in their functional cousins. Companies are working hard to address these shortcomings, primarily focusing on medically important regions that are difficult to sequence.

2. *The published pharmacogenomics literature has focused on common variation*. Most pharmacogenomic knowledge is related to variations occurring with a minor allele frequency of at least 5–10%. Initial studies focused on small cohorts for which high allele frequencies were required for statistical power. Genotyping arrays use common variations as tags for discovery of pharmacogenomics associations and typically focus on European populations. By contrast, genome and exome sequencing yield variations that are novel or rare. These occur in the midst of genes with known pharmacogenetics importance (or their associated pathways) and could have large effects. The impact of these variations can sometimes be estimated with computational algorithms, and larger data sets are providing better information about the phenotypic consequences of these variations.

3. *The published pharmacogenomics literature does not always identify risk alleles unambiguously*. Many papers do not report the chromosomal location or strand of a variant in an unambiguous manner and sometimes do not clearly specify the "risk" allele. Sometimes authors refer to the "common allele," but this is population dependent and therefore a treacherous convention. The PharmGKB curators work hard to resolve ambiguities and provide associations using an unambiguous standard.

4. *There are few pharmacogenomic effects that are associated with clear guidelines for clinical implementation*. The Clinical Pharmacogenomics Implementation Consortium has created guidelines for drugs with a strong genetic literature, including warfarin, codeine, clopidogrel, thiopurines, abacavir, simvastatin, and allopurinol. However, there is information to support the clinical implementation of pharmacogenomics for many other drugs. The PharmGKB is working on creating brief recommendations based on the literature, as a prelude to fully documented guidelines.

5. *We do not know how to combine the effects of multiple variants*. For some drugs, there are multiple variations known to impact response. An individual may have some alleles that suggest increased efficacy and others that suggest decreased efficacy. Only in rare cases have these combinations been assessed, and so it not clear how to combine this information. Large-cohort studies should provide sufficient data to untangle these interactions and should suggest principles by which the effects of multiple variations can be estimated.

6. *We do not know how to combine the effects of genetic variation on multiple drugs*. Even if we can integrate the impact of multiple variations on the response to a single drug, we must also assess their impact in the context of multiple drugs. Drug interactions often occur in the context of shared pathways of metabolism, and there are multiple ways in which other drugs can combine with genetic variation to modulate these pathways. The emerging field of systems pharmacology provides hope for quantitative models of drug action that will allow these multiple effects to be integrated.

7. *Phasing of the genome (computing haplotypes) is critical for interpretation*. To interpret multiple variants, we often must know whether these variants are on the same chromosome. For example, the father in the family we analyzed has two variations in *CYP2C19*—a truncation single-nucleotide polymorphism that renders one allele nonfunctional and a promoter single-nucleotide polymorphism that increases expression of one allele. The anticipated phenotype is quite different if the two single-nucleotide polymorphisms are on the same chromosome vs. different chromosomes. Fortunately, methods for both computational and experimental determination of haplotypes have emerged.

8. *We are just beginning to understand the impact of regulatory variation on drug response*. Our analyses were published before the publication of the ENCODE project. Variations in the regulation of genes can be important for drug response. There are already well-documented cases, such as the promoter polymorphism in *UGT1A* impacting response to irinotecan and other drugs. There are undoubtedly functionally relevant variations in noncoding DNA, and the ENCODE data sets provide a great opportunity to interpret variations in noncoding regions of the genome. In addition to these limitations in the technical capabilities and methods for interpretation and integration, we also observed the response of the individuals whose genomes were being analyzed (all consented to the sequencing and received genetic counseling).

9. *Individuals with annotated genomes discount the value of information about drugs they are not taking*. Although we were able to create limited clinical advice for nearly 100 drugs, study subjects were primarily interested in drugs that they were taking already or expected to take in the near future. They focused on “bad news” about drugs that would be less efficacious, more toxic, or more difficult to dose. In general, their responses highlighted the differences in presenting pharmacogenomic interpretations directly to consumers vs. providing them to medical professionals.

10. *Individuals with annotated genomes are frustrated by our inability to combine multiple variants with contradictory associations*. As noted above, we do not have general methods for combining variant effects into an overall predicted phenotype. Individuals are particularly concerned about this inability and interpret it as a major limitation of the field and the utility of the resulting analysis.

These observations provide a challenging set of research and implementation priorities for the next generation of pharmacogenomics. Narrowly defined, pharmacogenomics only estimates the genetic influence on drug response, but we know that there are many other critical influences, including compliance, other medications, diet, disease state, and other environmental exposures. The future for pharmacogenomics has two dominant themes. First, the technical task of accurate sequencing and accurate estimation of the influence of genetic variation on drug response. Second, the combination of this information with nongenetic predictors of drug response into an overall approach that supports optimal drug use, maximizing efficacy and minimizing toxicity.

Acknowledgments

PharmGKB is supported by National Institutes of Health grant GM061374.

References

1. Whirl-Carrillo M, et al. Pharmacogenomics knowledge for personalized medicine. Clin. Pharmacol. Ther. 2012; 92:414–417. [PubMed: 22992668]
2. Swen JJ, et al. Pharmacogenetics: from bench to byte–an update of guidelines. Clin. Pharmacol. Ther. 2011; 89:662–673. [PubMed: 21412232]
3. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clin. Pharmacol. Ther. 2011; 89:464–467. [PubMed: 21270786]
4. Ashley EA, et al. Clinical assessment incorporating a personal genome. Lancet. 2010; 375:1525–1535. [PubMed: 20435227]
5. Dewey FE, et al. Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. PLoS Genet. 2011; 7:e1002280. [PubMed: 21935354]
6. Chen R, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell. 2012; 148:1293–1307. [PubMed: 22424236]