An “Omics” View of Drug Development

Russ B. Altman,1,2 Daniel L. Rubin,2 and Teri E. Klein1,2

1Department of Genetics, Stanford University Stanford, California
2Stanford Medical Informatics, Stanford University Stanford, California

ABSTRACT

The pharmaceutical industry cannot be blamed for having a love/hate relationship with the fields of pharmacogenetics and pharmacogenomics. At the same time that pharmacogenetics and pharmacogenomics promise to save pipeline drugs by identifying subsets of the population for which they work best, they also threaten to increase the complexity of new drug applications, fragment markets, and create uncertainty for prescribers who simply do not understand or have time to master “personalized medicine.” Most importantly, the logical case for genetics-specific drug selection and dosing is much more mature than the practical list of drugs for which outcomes are demonstrably improved. Understandably, pharmaceutical developers and regulators have been careful in creating strategies for using genetics in drug development, and only recently has the FDA begun to establish preliminary rules for pharmacogenetic testing. A growing public academic effort in pharmacogenetics and pharmacogenomics is helping flesh out the basic science underpinnings of the field, and this should combine with extensive efforts of industry to create a solid foundation for future use of genetics in drug development. Two grand challenges to accelerate our capabilities include the characterization of all human genes involved in the basic pharmacokinetics of drugs, and the detailed study of the genes and pathways associated with G-protein-coupled receptors and how they are affected by genetic variation. Drug Dev. Res. 62:81–85, 2004. © 2004 Wiley-Liss, Inc.

KEY words: pharmacogenomics; pharmacogenetics; pharmacokinetics; GPCRs

INTRODUCTION

The term “pharmacogenetics” has traditionally been used to describe studies of a single gene and its dominant effect on the actions of a single (or small class) of drugs. In some ways, pharmacogenetics is the analogue to Mendelian genetics for inheritance: the variations of single genes translate in a straightforward way to phenotypic differences. A classic example is the metabolism of 6-mercaptopurine (6-MP) by thiopurinemethyltransferase (TPMT). Patients with certain variations do not metabolize 6-MP rapidly, and thus require careful (low) dosing of the medicine compared to patients with the most common version of TPMT [Weinshilboum, 2001]. The term “pharmacogenomics” is more recent, and has come to mean the study of gene-drug interactions on a larger scale and in a more comprehensive manner, usually using the tools of functional genomics (high throughput sequencing, expression analysis, and protein characterization). Pharmacogenomics is analogous to complex genetic inheritance, where single gene interactions are not sufficient to explain phenotypes, but entire systems...
must be studied completely to understand drug response. In general the "-omics" suffix has been used to denote the study of the entire set of entities in a class. Genomics studies all genes, proteomics studies all proteins, metabolomics studies all metabolites. By studying drug response from a systems perspective, it seems less likely that important gene-drug interactions will be missed, and thus the process of drug development may be made more efficient. For the remainder of this report, we will use the more general term, pharmacogenomics.

What, then, are "pharmacogenes" and why should we study all of them? In general, they can be defined as all the genes in the human genome that are important for determining/understanding/predicting the response to drugs. In general, pharmacogenes are either important for drug metabolism (pharmacokinetics, PK) or drug action (pharmacodynamics, PD). There is no gold-standard for determining if a gene is a pharmacogene, and splitters will say that virtually all genes are potentially important for drug response, and splitters will insist on evidence before declaring a pharmacogene. The PharmGKB database (http://www.pharmgkb.org/) maintains a curated list of published gene-drug relationships [Klein and Altman, 2004]. The curators of the database scan the pharmacogenomics literature and create and classify entries that link human genes to the metabolism or action of drugs. This collection currently lists associations for 1,349 genes with approximately 300 drugs, and so the splitters might choose 1,349 as the current best estimate of the number of pharmacogenes. A list of the most frequently mentioned drugs and genes in the context of pharmacogenomics is shown in Table 1.

One of the major difficulties for basic research in pharmacogenetics and pharmacogenomics is the management of large amounts of information, even about the basic players, in gene-drug-disease interactions. There are roughly 30,000 human genes, more than 5,000 drugs on the market, more than 4,000 diseases in the Medical Subject Headings, and uncountable numbers of measurable phenotypes (laboratory measurements, clinical measurements, symptoms, side effects, and signs) related to drug response. One recent study estimates that the PubMed database of medical literature contains roughly 200,000 articles that relate genes to drugs in one way or another (from a total of more than twelve million citations) [Rubin, and Altman, 2004]. Thus, there is a clear need for support in (1) assessing existing knowledge and (2) generating new hypotheses.

A CHALLENGE FOR PHARMACOKINETICS

The existing literature on pharmacogenomics is dominated by genes involved in the metabolism of drugs, including the major genes involved in the cytochrome p450 system: CYP2D6, CYP3A4, CYP2C19, CYP2C9, CYP1A2, and CYP3A5. These genes are significant for the pharmacokinetics of many drugs, and there are online clinical references that compile the association of clinical drug interactions with shared metabolism by these and other genes (see, for example, http://medicine.iupui.edu/flockhart/table.htm). Other genes that are on the list of "usual" pharmacokinetics suspects include the conjugating enzymes (acetylases, uridylnases, methyltransferases, sulfotransferases) and transporter genes (particularly, MDR1, organic anion, and cation transporters). The origin of these lists is partly historical. Once important drug effects were discovered, scientists began to check the effects of these same genes on other drugs, and the evidence for their importance increased. The degree to which there are important but undiscovered additional genes in these or other families is not entirely clear. However, one of the values of having the human genome sequence is that the complete set of genes is known. The existence of homologs to well-known drug-metabolizing genes allows us to systematically evaluate the presence and importance of their gene products. It would not be surprising to discover new variations (either homologs or post-translationally modified products) that have important effects. The key point, however, is that we have the tools necessary to identify and characterize the complete set of genes (cytochromes, conjugation enzymes, and transporters) that are likely to be important for the metabolism of all drugs. Gene sequence analysis gives us the candidate genes in the genome. Microarray expression experi-

TABLE 1. Most frequently mentioned drugs and genes in an automated scan for pharmacogenomics-related articles in the medical literature (PubMed). The first column lists the drug names, the second column lists the number of articles associating the drug with a gene (possibly with many duplicates) in PubMed, based on an analysis to be fully described elsewhere. The third column lists the gene names, and the final column is the number of articles associating the gene with a drug in PubMed [3].

<table>
<thead>
<tr>
<th>Drug</th>
<th># Articles</th>
<th>Gene</th>
<th>#Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mephenytoin</td>
<td>1349</td>
<td>CYP2D6</td>
<td>2601</td>
</tr>
<tr>
<td>Testosterone</td>
<td>1188</td>
<td>CYP3A4</td>
<td>1802</td>
</tr>
<tr>
<td>Phenobarbital</td>
<td>1137</td>
<td>CYP1A2</td>
<td>1470</td>
</tr>
<tr>
<td>Debrisoquine</td>
<td>1067</td>
<td>CYP2C19</td>
<td>1302</td>
</tr>
<tr>
<td>Quinidine</td>
<td>957</td>
<td>CYP2C9</td>
<td>1218</td>
</tr>
<tr>
<td>Ketoconazole</td>
<td>910</td>
<td>CYP3A</td>
<td>1206</td>
</tr>
<tr>
<td>Tolbutamide</td>
<td>795</td>
<td>CYP</td>
<td>841</td>
</tr>
<tr>
<td>Insulin</td>
<td>736</td>
<td>angiotensin</td>
<td>767</td>
</tr>
<tr>
<td>Troleandomycin</td>
<td>723</td>
<td>CYP2E1</td>
<td>702</td>
</tr>
<tr>
<td>Dopamine</td>
<td>v645</td>
<td>renin</td>
<td>479</td>
</tr>
</tbody>
</table>
ments allow us to evaluate mRNA expression patterns in different tissues. Proteomics (mass spectrometry) should allow characterization of the protein products and their further localization.

Thus, one grand challenge for pharmacogenomics could be the complete characterization of all genes available for the metabolism of drugs, their gene expression, protein translation, modifications, and localization. Although variation in these genes represents only half the story for pharmacogenomics (the PK half), it is a half that seems to be within reach with the completion of the human genome, and the concomitant ability to identify all the enzymes and transporters present in the genome. The implications of such a complete catalog for drug development would be significant. With a finite list of genes relevant to pharmacokinetics, new drugs could be evaluated with respect to their likely interactions with gene products. Instead of waiting to learn about untoward interactions in clinical studies, high throughput experimental or computational assays could be used to gauge drug affinity to or metabolism by allelic variants of enzymes and transporters. The chances of surprise effects (at least for PK) would be reduced because all potential interactions could be screened.

**A CHALLENGE FOR PHARMACODYNAMICS OF GPCRS**

The impact of genetic variation on drug response and pharmacodynamics is harder to assess. While metabolism seems to be carried out by a relatively well-defined and identifiable subset of genes, drug action can occur on a much larger and more diverse set of target genes. Thus, the chances that new target genes will be fully characterized in terms of their functional characteristics are lower. However, recent surveys of prescription drugs indicate that G-protein coupled receptors (GPCRs), taken as a group, are the most common set of targets. These include receptors for histamine (allergy and antacid medications), angiotensin (hypertension), α and β adrenoceptors (asthma, hypertension, cardiac), insulin (diabetes), serotonin (depression), and many others. Although the drugs that act on these receptors are quite diverse, they can be categorized based on how they interact with their receptor (activator or inhibitors, for example). There is an increasing pace of publication in the public domain on the genetic variations in these receptors and how they relate to phenotypes. For example, variations in the beta-adrenergic protein sequence [Green et al., 2001] as well as promoter regions [Drysdale et al., 2000] have been shown to correlate with variation in drug response. It may be early to summarize the overall lessons on this group of genes, but therein may lie an opportunity for systematic study of the ways in which genetic variation affects response.

A second grand challenge for pharmacogenomics, then, may be the characterization of the structure and function of GPCR genes and protein products in such a way that important variations in these can be associated with variation in response to the drugs that target them. These genes are generally homologs, and so the location of active sites, inhibitory sites, activating sites, and interaction sites are in some cases already mapped or actively being mapped. They are also often part of homologous signaling pathways, and so systematic study of these pathways may yield recurring motifs in how variation affects drug response. There are large-scale efforts to characterize cell signaling in a manner that is helping to address this challenge, but the focus has, so far, been chiefly on basic mechanisms and not on variations due to genetic polymorphism. The regulation of expression of these genes is of great interest, and the degree to which polymorphisms affecting expression will have shared features that go across the family is worth evaluating.

Aside from GPCRs, there are a large number of drug targets for which genetic variation remains to be assessed and catalogued. Fortunately, common variations in human genes only need be catalogued once, and so focused efforts on genes of pharmacologic interest (such as are undertaken in academic and industrial efforts focused on particular targets) as well as general surveys of variation (such as in the HapMap project) [Consortium, 2003] are likely to yield the fundamental list of common variations individually (SNPs) as well as common combinations of these variations (Haplotypes). The remaining challenge then, and one that is likely to persist for a while, is to characterize the functional consequences of these variations for drug response. High throughput genomic technologies will be critical for characterizing these functional consequences. The ability of mRNA expression arrays to simultaneously measure the mRNA levels of all genes in a particular cell type helps ensure that the important consequences of drug exposure on gene expression are not missed. Similarly, proteomic technologies should be developed to assay levels and activities of gene products in a comprehensive, cost-effective manner. Until or unless common themes emerge that are repeated across many targets, it is likely that understanding the pharmacogenomics of many drug-target interactions will be a one-at-a-time activity, undertaken by laboratories with a particular interest (industrial or academic) in these targets or their associated drugs. The amount of work to be done is considerable, and so the real issue becomes prioritization of effort.
IDENTIFYING RESEARCH PRIORITIES

Industrial and academic efforts in pharmacogenomics share the same basic science, but are driven by different fundamental interests. These affect how they prioritize scientific opportunities in pharmacogenomics. The fundamental interest in pharmacogenomics for industry is the creation of new pharmaceutical products for which the public will pay (in order to improve health). The fundamental interest in academic pharmacogenomics is the creation and dissemination of new knowledge about the principles by which human genetic variation leads to variation in drug response. There are two main areas in which these different interests manifest themselves as different research programs: the choice of drug systems and the willingness to focus on rare variations. The pharmaceutical industry is interested in high-impact drugs and diseases that are of interest to a large fraction of the population. Thus, they will naturally focus on drugs for common diseases. They will also focus generally on genetic variations that are found commonly in the population and thus would impact many patients. A caveat to the “looks at common variations” rule would be cases when a rare variation is associated with catastrophic outcomes and thus threatens the market viability of a drug, which is otherwise safe for most patients. Academic research units have a different calculus for pharmacogenomics: drugs and diseases may be selected specifically to be rare (in order to avoid competition with pharmaceutical companies), but are certainly selected because they are amenable to detailed study, based on the availability of model systems, the relevance of previous results to set the stage for progress, the amenability of the system to study with novel experimental techniques, and the ability to validate and publish findings about the underlying basic biological principles. Similarly, rare variants that illustrate important principles of pharmacogenomics are welcomed by academics.

The differences in the pharmacogenomic research programs of academics and industrial scientists thus span the space of opportunities fairly well. The long-term impact of pharmacogenomics, however, will be based on whether we find genetic variation that is cost-effective to measure before prescribing a drug. There are three main scenarios for the cost-effectiveness of pharmacogenomic interventions. First, the scenario where outcomes such as life and death depend on the prescribing decision, and thus the value of getting the right dose of the right drug is very high for a particular individual. This scenario may be associated with the decision to prescribe a drug to treat a cancer or severe depression, where failure can be disastrous. Second is the scenario where differences in outcomes are more moderate, but occur so frequently that the net impact on health is significant. This scenario may be associated with treatments for diseases for which there is a heavy society burden (hypertension, diabetes, cardiac disease) where small increases in efficacy and/or small decreases in toxicity can have a large aggregate effect on outcomes such as length of life, quality of life, and days of work missed. The final scenario is that in which detailed knowledge of the biology of a rare disease allows drugs to be created at markedly reduced overall costs, and thus makes new treatments available in settings where traditional drug development would not be cost effective. This scenario may be associated with rare diseases (perhaps studied only in an academic setting) for which high throughput functional genomic methods allow identification of compounds with a high likelihood of clinical success. This final scenario may be the most wishful, but detailed understanding of major pathways associated with common diseases may reasonably be expected to yield “spin off” knowledge relevant to more rare diseases.

The success of other “omics” initiatives has led to a new research ethic: devise relatively inexpensive high-throughput experiments to collect comprehensive data sets and store these data in databases with appropriate search and comparison functionalities. These databases are then used to evaluate opportunities for new knowledge by mining the databases to form specific testable hypotheses and subsequently returning to the lab for confirmation. The advantage of this approach should be twofold. First, the use of comprehensive data sets collected without particular hypotheses in mind should lead to a less biased, more objective search for promising hypotheses. Second, these data sets should allow the resulting hypotheses to be very focused and straightforward to test with confirmatory experiments, thus increasing the overall cost-effectiveness of the scientific enterprise. The pharmacology research community has bet that an —omics approach to drug discovery and development will result in effective use of genomic information to accelerate our ability to prescribe the right drug at the right dose on the first try.

ACKNOWLEDGMENTS

This work is supported by NIH GM61374, which funds the PharmGKB (http://www.pharmgkb.org/).

REFERENCES


