# Genetic Variation as a Guide to Drug Development

Patrick W. Kleyn* and Elliot S. Vesell

Genomic technologies have increasingly been adopted in the biotechnology and pharmaceutical industries to accelerate identification of disease-validated targets for drug discovery. When used with combinatorial chemistry and high-throughput screening techniques, it is likely that genomics will identify more novel chemical entities (NCEs) for drug development. Historically, more than 90% of NCEs entering preclinical development never reach the clinic (*1*). Many failed compounds have unacceptable toxicity profiles in animal studies or initial testing in humans, whereas others show insufficient efficacy. Even after clinical trials involving hundreds of patients, there is still the risk of unexpected toxicity in a subset of the population such as occurred with fenfluramine (Redux), terfenadine (Seldane), and, most recently, mibefradil (Posicor).

TECHVIEW
PHARMACOGENOMICS
Mailbox: www.sciencemag.org/dmail.cgi?53841

Pharmacogenetics is the study of genetic variation underlying differential response to drugs. In the table are examples of pharmacogenetic markers that were discovered through a hypothesis-driven approach in which each polymorphism was identified in a plausible candidate gene. Pharmacogenomics applies the large-scale systematic approaches of genomics to speed the discovery of drug response markers, whether they act at the level of the drug target, drug metabolism, or disease pathways.

Pharmacogenetic studies have established the importance of polymorphic drug-metabolizing enzymes (DMEs) such as CYP2D6, a member of the cytochrome P450 superfamily, in the differential response of patients to drugs. Only more recently have genetic factors at the level of the drug target or the disease pathway been identified. For example, *ApoE4*, an allele at the apolipoprotein E locus, not only correlates with an elevated risk of developing Alzheimer's disease (AD), but also predicts poor response to cholinesterase inhibitor treatment (*2–4*). This is an example of a polymorphism within a disease-related gene that is predictive of drug response. It seems likely that for many common diseases including cancer, atherosclerosis, and the neurodegenerative disorders each represents a collection of separate conditions with a similar clinical endpoint, but they have distinct etiologies and therefore distinct responses to therapy. That is the underlying hypothesis of pharmacogenomics. Below, we examine the technologies that are enabling pharmacogenomic studies.

The methodology of genome-wide DNA genotyping as applied to pharmacogenomic studies evolved from linkage and association studies of complex diseases. Linkage studies involve genotyping families with microsatellite markers (variable numbers of short tandem repeats). The goal is to correlate inheritance of a particular chromosomal region with inheritance of disease. Within a family, shared regions are typically tens of millions of bases long. Thus, only a few hundred microsatellite markers evenly spaced throughout the genome suffice to detect any particular region. However, because drug response data can rarely be obtained from multiple members of a family, linkage studies are impractical for pharmacogenomics. Association studies, by contrast, correlate the presence of a chromosomal region and a trait (disease or drug response) in unrelated individuals of a population. Because the common ancestry of unrelated individuals in an open population is much more distant than that of family members, the shared chromosomal regions are typically much smaller, 100 kilobases or less. Thus, in order to perform association studies on a genome-wide basis in an open population, 100,000 markers or more are required. Such dense maps are not yet available. However, the rapid pace of DNA marker discovery together with novel genotyping technologies will soon permit genome-wide association studies.

Whereas a large linkage study requires about 500,000 genotypes (1000 family members with 500 microsatellite markers), a genome-wide association study might well necessitate 100 million genotypes (1000 individuals typed for 100,000 markers). These technical considerations favor use of single nucleotide polymorphisms (SNPs) rather than microsatellite markers used for linkage studies. SNPs are simple base-pair substitutions that occur within and outside genes. Although biallelic SNPs are inherently less polymorphic than multiallelic microsatellites, they occur much more frequently [roughly every 500 base pairs (*5*)]. They are also less mutable and more amenable to automated genotyping.

The rapid accumulation of human genomic and mRNA sequences over the next few years will lead to the identification of thousands of SNPs for genome-wide association studies. However, there are still potential hurdles and limits to this approach. One confounding factor will be allelic heterogeneity. In open populations, multiple alleles in several genes often underlie the heritable component of a trait. In a linkage study, all alleles at a particular locus contribute to signal; in contrast, an association study might detect the contribution of only a single allele. Consequently, associations are detected with nearby markers in diseases with common alleles such as cystic fibrosis (*Δ508*) or Alzheimer's disease (*ApoE4*). However, genes with many rare alleles such as lipoprotein lipase (*LPL*) may not be detected by genome-wide association studies (*5*). Ultimately, the success of this approach will depend on the population-specific relative risk associated with each drug response allele. Alleles with a strong relative risk such as *ApoE4* for AD in Caucasian populations can be detected in genome-wide studies, as recently demonstrated by Zubenko *et al.* (*6*). Using 391 microsatellite markers spaced throughout the genome, they compared 50 AD patients with 50 controls. They detected significant association with six markers, including one located within 2.4 centimorgans of *ApoE*. However, alleles with weaker effects such as those within the cholesteryl ester transfer protein (*CETP*), β-*fibrinogen*, and *LPL* genes, which predict increased response to the 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitor pravastatin, may prove harder to detect in genome-wide association studies (*7–9*).

The phenotypic consequences of polymorphisms affecting drug response are presumably associated with changes at the RNA and protein level. Transcriptional analyses thus provide another approach for identifying pharmacogenomic markers. Large-scale transcriptional studies are possible primarily as a result of the development of dense cDNA and oligonucleotide arrays that are hybridized to fluorescently

P. W. Kleyn is at Millennium Predictive Medicine, Inc., 640 Memorial Drive, Cambridge, MA 02139–4815, USA. E. S. Vesell is at the Department of Pharmacology, Pennsylvania State University College of Medicine, Hershey, PA 17033, USA.

*To whom correspondence should be addressed. E-mail: kleyn@mpi.com.

**EXAMPLES OF PHARMACOGENETIC MARKERS BY MECHANISM**

| Mechanism & Genes | Pharmacogenetic effect |
|---|---|
| **Drug metabolism** | |
| *N*-Acetyltransferases (*NAT1* & *NAT2*) | The "slow acetylator" phenotype (40 to 60% of Caucasians) arises from 3 major alleles in *NAT2* and results in slow clearance and associated toxicity of multiple drugs including isoniazid, phenelzine, and procainamide (*16*). |
| *CYP2D6* | *CYP2D6* defect leads to "poor metabolizer" (PM) phenotype in 5 to 10% of Caucasians. Metabolizes ~25% of all drugs including many cardiovascular drugs and antipsychotics (*17*). |
| **Drug target** | |
| 5-$HT_{2A}$ receptor | Polymorphisms in the coding region and promoter of this serotonin receptor have been associated with efficacy of the atypical antipsychotic, clozapine (*18*). |
| Thymidylate synthase (*TS*) & dihydrofolate reductase (*DHFR*) | Overexpression of *TS* and *DHFR* are mechanisms by which tumor cells acquire resistance to antimetabolites such as 5-fluorouracil and methotrexate (*19, 20*). |
| **Disease pathway** | |
| Apolipoprotein E | *ApoE4* allele is associated with risk of developing Alzheimer's disease (AD) as well as response to tacrine treatment of AD (*3*). |
| Cholesteryl ester transport protein (*CETP*), lipoprotein lipase (*LDL*), and β-fibrinogen | Polymorphisms in all 3 genes have been associated with atherosclerosis progression as well as response to the HMG-CoA reductase inhibitor, pravastatin (*11–13*). |

or radioactively labeled RNA or cDNA. By comparing hybridization signals from RNAs extracted from two different cell populations, differentially regulated genes can be identified.

Pharmacogenomic applications of array-based transcript profiling include analysis of patient tissues in response to therapy during clinical trials. Expression-based studies appear to be especially appropriate in cancers, because RNA can be obtained from biopsies and surgical specimens. This technology readily detects the somatic changes associated with the development of some tumors and their response to chemotherapy. Somatic changes linked to therapeutic outcomes include the amplification of the oncogene *erb-B2*, which predicts good response to cyclophosphamide–methotrexate–5-fluorouracil (CMF) adjuvant therapy of breast cancer (*10*). In other diseases, transcript profiling can be used to identify candidate genes. For example, transcript profiling of epithelial and inflammatory cells isolated from bronchial lavages might identify candidate markers for response to asthma therapeutics. These candidate genes could then be screened for polymorphisms, which could be typed in a larger population of patients. Alternatively, expression of genes that predict drug response might be routinely assayed at the protein level by antibody-based tests in sputum or serum.

Current areas of technology development in transcript profiling include RNA amplification protocols that permit use of low quantities of starting material (*11*); laser capture microdissection (LCM) (*12*), which facilitates isolation of individual cells from contaminating material in heterogeneous clinical samples; and continuing development of arrays and associated imaging systems to improve sensitivity. Without highly sensitive arrays, subtractive cloning protocols that enrich differentially expressed transcripts play an important role in profiling heterogeneous clinical samples (*13*). Because subtractive cloning is PCR-based, it is particularly valuable for detecting low-abundance transcripts.

For many pharmacogenomic and diagnostic applications, proteins remain markers of choice. The relative stability of proteins in clinical specimens and the use of simple enzyme-linked immunosorbent assays (ELISA) on high-throughput platforms allow implementation of these markers. Large-scale protein analysis of clinical samples can be carried out by two-dimensional separation by electrophoresis on polyacrylamide gels (2D-PAGE) coupled with mass spectrometry for protein identification. However, the complexity of clinical specimens and the limited reproducibility, throughput, and sensitivity of 2D-PAGE have hampered wide application of this approach.

However, recent improvements in mass spectrometer sensitivity and throughput together with continuing developments in 2D-PAGE and chromatographic separation suggest the possibility of "proteomic" analysis akin to genomic studies [reviewed in (*14, 15*)]. Proteomics may soon emerge as a powerful approach for directly identifying highly predictive pharmacogenomic markers in blood or other bodily fluids.

In the short term, pharmacogenomics will be strategically used for clinical development of particular compounds with potential efficacy or toxicity issues. Pharmacogenomics may also be applied to approved drugs with restricted market share because of limited efficacy or high toxicity. As our molecular understanding of both human populations and preclinical models of disease and toxicity develops, pharmacogenomic information will optimize predictions of drug effects in humans.

Skillful physicians have always recognized individual variations in disease expression and drug response. The art of medicine consists in identifying and successfully overcoming these challenges. Pharmacogenetics was the first attempt to establish a scientific basis for these variations. Pharmacogenomics carries on that tradition and broadens the inquiry by enlisting modern technologies described in this review.

**References**


1. J. Drews, *Nature Biotechnol.* **14**, 1516 (1996).
2. E. H. Corder *et al.*, *Science* **261**, 921 (1993).
3. J. Poirier *et al.*, *Lancet* **342**, 697 (1993).
4. J. Poirier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 12260 (1995).
5. D. A. Nickerson *et al.*, *Nature Genet.* **19**, 233 (1998).
6. G. S. Zubenko *et al.*, *Genomics* **50**, 121 (1998).
7. M. P. M. de Maat *et al.*, *Arterioscler. Thromb. Vasc. Biol.* **18**, 265 (1998).
8. J. W. Jukema *et al.*, *Circulation* **94**, 1913 (1996).
9. J. A. Kuivenhoven *et al.*, *N. Engl. J. Med.* **338**, 86 (1998).
10. H. B. Muss *et al.*, *ibid.* **330**, 1260 (1994).
11. R. N. Van Gelder *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**,1663 (1990).
12. M. R. Emmert-Buck *et al.*, *Science* **274**, 998 (1996).
13. L. Diatchecko *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6025 (1996).
14. M. R. Wilkins *et al.*, Eds., *Proteome Research: New Frontiers in Functional Genomics* (Springer-Verlag, New York, 1997).
15. A. Persidis, *Nature Biotechnol.* **16**, 393 (1998).
16. K. P. Vatsis, K. J. Martell, W. W. Weber, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6333 (1991).
17. J. R. Idle and R. L. Smith, *Drug Metab. Rev.* **9**, 301 (1979).
18. M. J. Arrantz *et al.*, *Neurosci. Lett.* **217**, 177 (1996).
19. M. Kornmann *et al.*, *Cancer Lett.* **118**, 29 (1997).
20. J. M. Trent *et al.*, *J. Clin. Oncol.* **2**, 8 (1984).