

Indexing pharmacogenetic knowledge on the World Wide Web

Russ B. Altman^a, David A. Flockhart^b, Stephen T. Sherry^c, Diane E. Oliver^a, Daniel L. Rubin^a and Teri E. Klein^a

Pharmacogenetics 2003, 13:3–5

^aDepartment of Genetics, Stanford Medical Informatics, Stanford University, Stanford, California, ^bDivision of Clinical Pharmacology, Indiana University, Indianapolis, Indiana and ^cthe National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland, USA.

A key challenge for pharmacogenetics is the creation of databases to store, analyse and disseminate important datasets in order to catalyse research and training. Most successful databases have a limited scope: Genbank contains DNA sequences [1]; the Protein Data Bank contains the three-dimensional coordinates of macromolecules [2]; the Online Mendelian Inheritance in Man contains a record of human genetic disease [3]; and PubMed contains the biomedical literature [4]. This limited scope is a great strength, because the information can be stored, searched and analysed using a few powerful tools, and the users of these databases know exactly what to expect. Databases for pharmacogenetics and pharmacogenomics will have much more diversity. Pharmacogenetic data involve phenotypes that are as diverse as the assays we invent to measure them. Thus, it is unclear what a user should expect from a pharmacogenetics database, and yet a public repository of pharmacogenetic data is critical to establish a core dataset for the field upon which we can build new analyses and new hypotheses [1]. Clearly, successful databases for pharmacogenetics must employ some sort of classification of phenotypes that is general purpose, yet extensible to include undefined characterizations of phenotype.

We propose a classification to organize pharmacogenetic information based on two principles. First, pharmacogenetics relates variation in a *gene* to variation in some *phenotype* associated with a *drug*, and these three concepts should be used as major labels for most pharmacogenetics datasets. Second, we can further classify phenotypic data based on the nature of the data collected. In particular, we propose five types of variation information that commonly are reported in journals such as *Pharmacogenetics*, and which provide reasonable coverage of the field (Fig. 1).

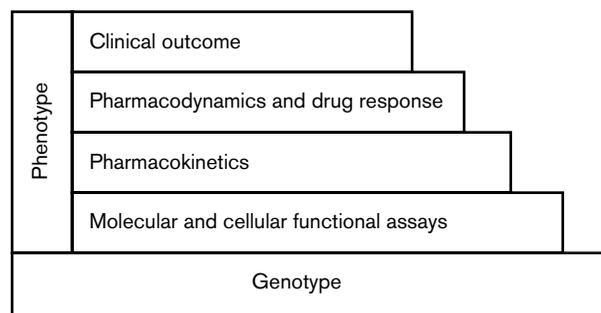
Genetic variation

The most basic information about pharmacogenetic variation is the observation of variation in individual genes, the type and location of the variation, and the frequency of the variation in the populations of interest. Thus, the first type of pharmacogenetic variation information is genetic variation itself (this is actually independent of individual drugs, unlike the other types). The well described polymorphisms of *CYP2D6* (*4, *7, *8, *3, *6) are prototypes [5].

Molecular and cellular assay variation

Frequently, reports in the pharmacogenetics literature contain data establishing an association between genetic variation and the results of drug-related molecular or cellular assays. For example, we might find that the specific activity of an enzyme is reduced by a cSNP (molecular variation) or that the expression of a set of

Fig. 1



Variation in genotype and phenotype in the context of drugs is the central dogma of pharmacogenetics. We propose five categories of information where variation can be observed in the context of drugs that form a basis for indexing important pharmacogenetic datasets.

genes is altered by a cSNP in the presence of a drug (cellular variation). The change in 6-thioguanine nucleotide concentrations in the presence of TPMT variants is an example of this type of variation [6].

Pharmacokinetic variation

The early days of pharmacogenetics relied upon observing changes in drug metabolism in the context of genetic variants. Any genes that are involved in the absorption, distribution, metabolism or elimination of a drug might result in pharmacokinetic changes. Many datasets therefore demonstrate that genetic polymorphisms lead to variation in the pharmacokinetics of particular drugs. The well documented affect of CYP2C9 polymorphisms on warfarin levels is a well known example [7].

Pharmacodynamic and drug response variation

The action of a drug at its molecular receptors or targets can vary in response to genetic variation in these targets. Thus, many pharmacogenetic datasets document that the pharmacodynamic response to a drug (usually at the whole-organism level) vary as one or a set of genes vary. For example, we may find that polymorphisms in an opioid receptor lead to changes in pupillary constriction [8].

Outcomes variation

The pharmacogenetics community is acutely aware that, in order to make the impact upon clinical medicine, genetic variations in the response to drugs must have measurable differences in the clinical outcomes that patients and their physicians care about: drug side-effects, disability, days-of-work-missed, pain and death. Datasets that associate genetic variation with variation in clinical outcome are different from pharmacodynamics and drug response datasets, which may show differences that are not sufficient to alter practice or policy. An example of variation in outcomes is found in the data on the cure rates for peptic ulcer disease in patients treated with omeprazole with different variants of CYP2C19 [9].

These five areas are admittedly coarse, and there will certainly be datasets that touch upon more than one area, and should be multiply indexed. In addition, refinements in this high-level classification scheme could be made to create a tree of indexing terms. One value of these distinctions is that different types of pharmacogenetic researchers will contribute different types of datasets, based on their interests and capabilities. Thus, geneticists and genomics scientists will generate datasets of the first type, molecular and cellular biologists will generate datasets of the second type, pharmacologists will generate datasets of the third and fourth type, and clinical researchers and outcomes

specialists will generate outcomes datasets. The emergence of pharmacogenomics may blur the line between these specialties, and create new inter-disciplines. Thus, the indexing scheme corresponds only to an approximate set of scientific partitions within the field, but it should catalyse interactions between researchers within them.

A web site to facilitate identification and indexing of datasets

We are collaborating in the construction of the first comprehensive pharmacogenetic knowledge base, the PharmGKB (<http://www.pharmgkb.org/>), an NIH-funded project to bring together datasets from all these levels, index them, and represent them in a way that allows the community to retrieve and analyse them. The PharmGKB is charged with collecting and organizing major publicly available datasets in pharmacogenetics and pharmacogenomics. The task of identifying and extracting datasets from the literature and from participating research groups is ongoing, and will lead to a critical mass of data for the community. However, in order to accelerate public participation and guidance in our efforts, we have created a web site where the pharmacogenetics community can specify gene–drug interactions where genetic variability has been observed and associated with phenotypic variability at one or more of the levels.

We call this the ‘PharmGKB Community Submission Project’ and it is meant to help us identify and catalog important datasets in pharmacogenetics. It is available at <http://www.pharmgkb.org/community/>. Users submit a gene, a drug, a category of knowledge (from the above list), and their email address. They are also encouraged to provide links to the literature or other databases that support these gene–drug variability associations (and that provide hints on how to track down the data), as well as comments, key words and other information. We will collect these submissions, store and organize them in an accessible database. We will also allow public comment on individual submissions, to stimulate debate.

We believe that there are numerous benefits to be gained by this activity. First, it takes advantage of the Internet to accelerate the acquisition of important datasets for PharmGKB. Second, it takes advantage of the scientific diversity in pharmacogenetics to ‘cover the waterfront’ of potential sources of important datasets. Third, it creates a forum for scientists to be introduced to the PharmGKB and to other scientists, through their submissions and discussions. Fourth, it creates a portal into the more detailed PharmGKB datasets, as ‘hits’ on searches of the community-based site will be linked to datasets that we have acquired. Finally, it creates a tool for identifying new research

opportunities in pharmacogenetics. For example, a gene–drug interaction may have good genetic polymorphism information, outcomes information and molecular/cellular phenotype datasets, but be relatively unexplored in the areas of pharmacokinetics or molecular pharmacodynamics. We are optimistic that this prototype project will be successful, but some potential drawbacks include lack of participation, careless data entry, poor user interface and inadequate curation.

If successful, the PharmGKB Community Submission Project will become part of the research infrastructure that accelerates our ability to improve the development and delivery of drugs to patients. It may also serve as a model for other genotype-to-phenotype efforts that involve the complex datasets and multiple scientific disciplines.

Researchers are encouraged to visit the site, and to inform us of important gene variations that may affect the response to drugs.

References

- 1 Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. GenBank. *Nucleic Acids Res* 2002; **1**:17–20.
- 2 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000; **28**:235–242.
- 3 Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2002; **1**:52–55.
- 4 National Library of Medicine. PubMed. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>.
- 5 Stamer UM, Bayerer B, Wolf S, Hoeft A, Stuber F. Rapid and reliable method for cytochrome P450 2D6 genotyping. *Clin Chem* 2002; **48**:1412–1417.
- 6 Andersen JB, Szumlanski C, Weinshilboum RM, Schmiegelow K. Pharmacokinetics, dose adjustments, and 6-mercaptopurine/methotrexate drug interactions in two patients with thiopurine methyltransferase deficiency. *Acta Paediatr* 1998; **87**:108–111.
- 7 Takahashi H, Echizen H. Pharmacogenetics of warfarin elimination and its clinical implications. *Clin Pharmacokinet* 2001; **40**:587–603.
- 8 Lotsch J, Skarke C, Grosch S, Darimont J, Schmidt H, Geisslinger G. The polymorphism A118G of the human mu-opioid receptor gene decreases the pupil constrictory effect of morphine-6-glucuronide but not that of morphine. *Pharmacogenetics* 2002; **12**:3–9.
- 9 Furuta T, Ohashi K, Kamata T, Takashima M, Kosuge K, Kawasaki T, *et al.* Effect of genetic differences in omeprazole metabolism on cure rates for *Helicobacter pylori* infection and peptic ulcer. *Ann Intern Med* 1998; **15**:1027–1030.