OMICS A Journal of Integrative Biology
Volume 10, Number 4, 2006
© Mary Ann Liebert, Inc.

Integrating Large-Scale Genotype and Phenotype Data

TINA HERNANDEZ-BOUSSARD, MARK WOON, TERI E. KLEIN, and RUSS B. ALTMAN

ABSTRACT

With the completion of the Human Genome Project, a new emphasis is focusing on the sequence variation and the resulting phenotype. The number of data available from genomic studies addressing this relationship is rapidly growing. In order to analyze these data as a whole, they need to be integrated, aggregated and annotated in a timely manner. The Pharmacogenetics and Pharmacogenomics Knowledge Base PharmGKB; (<www.pharmgkb.org>) assembles and disseminates these data and their associated metadata that are needed for unambiguous identification and replication. Assembling these data in a timely manner is challenging, and the scalability of these data produce major challenges for a knowledge base such as PharmGKB. However, it is only through rapid global meta-annotation of these data that we will understand the relationship between specific genotype(s) and the related phenotype. PharmGKB has confronted these challenges, and these experiences and solutions can benefit all genome communities.

INTRODUCTION

The completion of the Human Genome Project has set the stage for a successor in this post genomic era—the study of population variation and its association with complex traits and phenotypes. Advances in genomic technologies have catalyzed this era with the production of high-throughput genome-wide arrays used to manipulate and analyze genomic information. These arrays measure data across an entire genome, including both open reading frames and intergenic regions. As these technologies become affordable and widely available, they will greatly increase the volume of genomic data available in the public domain, measuring data across the genome of multiple individuals. The amount of data presented in a single assay raises new challenges in the postgenomic era.

PharmGKB (<www.pharmgkb.org>) is a knowledge base that curates genomic information related to drugs, diseases, and genes, with a particular emphasis on gene variation and gene products, including data generated from high-throughput genome-wide arrays (Klein and Altman, 2004). The majority of the primary data in PharmGKB are single nucleotide polymorphisms (SNP), their surrounding sequences, details of their assay, and any related phenotypic measured data. Value is added to these data with the integration of expression data, from both exon and gene microarrays and other comparative genomic information. PharmGKB is capable of handling volumes of genomic data, the main goal being to catalyze pharmacogenomics research.

Stanford University Medical School, Department of Genetics, Stanford, California.

Previously, the majority of the data in PharmGKB has been from low-throughput assays with a few variants genotyped in several individuals. PharmGKB has expanded both its database schema and data storage capacity to integrate large-scale high-throughput genomic data, including studies with large numbers of samples (several thousand) assayed across the entire genome (hundreds of thousands of markers). To our knowledge, PharmGKB is the first public database that allows one to query, browse, and download genetic variant data (i.e., individual SNPs from an array) and its related expression data (both gene and exon) from one integrated system. As PharmGKB captures these genomic data, it must address the challenges associated with the integration, aggregation, and annotation of large-scale genome-wide data in the public domain. These challenges include the scalability and efficiency of data integration and the variety of data types across all genomic domains, from the single genetic variant (SNP) to the gene product and all external data related to each level. Finally, PharmGKB addresses problems arising from annotating these data as a whole and capturing metadata from different resources. Our experiences and insight regarding these challenges of "just-in-time" integration of expression data with other genomic data to support research goals should be relevant to all genomic domains.

PharmGKB

The Pharmacogenomics Knowledge Base (PharmGKB) is a public database funded by the National Institutes of Health since 1999 through the PharmacoGenetics Research Network (PGRN) to aid researchers investigating how genetic variation affects drug response. PharmGKB is one of several centers funded by the PGRN, and groups within the PGRN deposit their primary data into PharmGKB. Focusing on the relationship between genetic variation and drug response, most primary data within PharmGKB has been either genotype or phenotype low-throughput data. Experiments that focus on pharmacodynamics and pharmacokinetics are of key interest and are captured in PharmGKB. Pharmacodynamic studies focus on genetic variation in drug targets, aiming to link measurable differences in drug response to a genetic variant at the whole-organism level. Pharmacokinetic studies look at the relationship between the concentration of drugs or their metabolites and genetic variation in processes involved in the absorption, distribution, metabolism, or elimination of a drug (Tate and Goldstein, 2004). Any dataset that has established a clear relationship between a drug and/or disease with a specific genetic component is therefore considered for inclusion in PharmGKB, both from PGRN and non-PGRN groups. These relations are also represented in pharmacogenomic pathways (Fig. 1), developed by PGRN members and PharmGKB. PharmGKB's interactive user interface for pathways is unique and designed to be easily understood through the use of shapes and colors to depict genes, drugs, etc. Genes and drugs are hot-linked to PharmGKB gene and drug web pages, making the user one click away from extra information regarding these objects.

These relations are also represented in pharmacogenomic pathways (Fig. 1) developed by PGRN members and PharmGKB. PharmGKB's interactive user interface for pathways is unique and designed to be easily understood through the use of shapes and colors to depict genes, drugs, and other objects. Genes and drugs are hot-linked to PharmGKB and drug web pages, making the user one click away from additional information regarding these objects.

As a portal to pharmacogenomics research, PharmGKB collects and stores phenotype data and genotype data using XML in a schema developed by our group specific for representation and validation of genotype data (Fig. 2). Data captured includes sequence information, location of variants and exons on the gene, alternative splicing, changes in amino acid sequence, haplotype data, subject information such as gender and ethnicity, methods to assay the variant and more than 100 other relevant data components, which are then rendered on the website (Fig. 3). In addition to these primary data, we also report related external resources to integrate with our primary data, such as UCSC's GoldenPath (Kent et al., 2002), dbSNP (Sherry et al., 2001), jSNP (Hirakawa et al., 2002), and HapMap (International-HapMap-Consortium, 2005). All these data are available on PharmGKB and can be downloaded in raw data files or via web services through automated queries (<www.pharmgkb.org/home/projects/webservices/>). The aggregation of these data al-

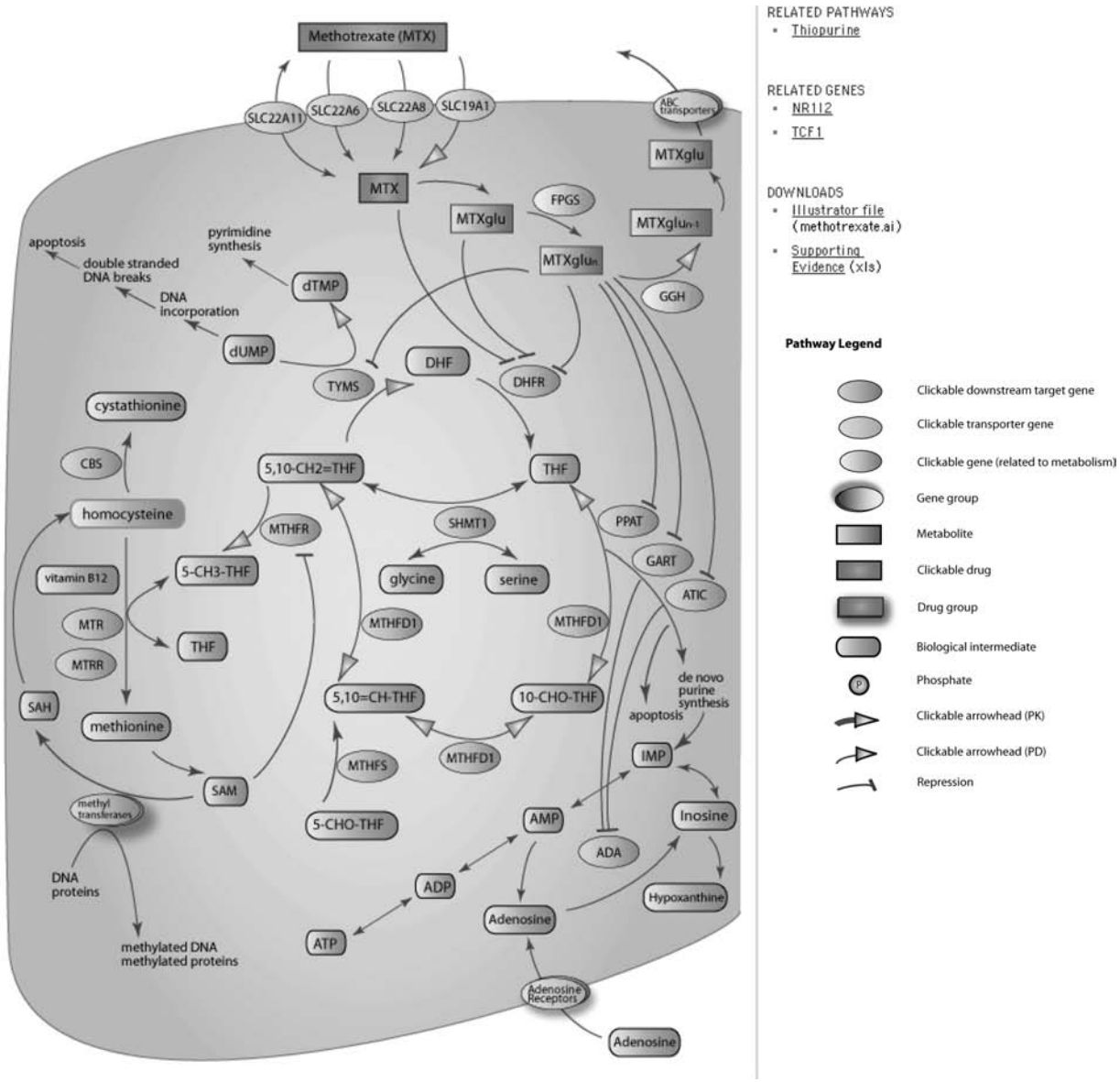


FIG. 1. The Methotrexate Pathway, illustrates the genes involved in mediating the effects of the antimetabolite methotrexate, an analogue of reduced folate, which targets endogeneous cellular folate metabolism. Drugs are depicted by purple boxes, transporter genes by turquoise ovals, and biological intermediates by green rounded boxes. Available at: http://www.pharmagkb.org/do/serve?objId=PA2039&objCls=Pathway

lows us to give functional meaning to individual variants that can aid in prediction of alternative product function, giving light to the underlying mechanisms of the variants.

PharmGKB has been used for several years for the distribution and exchange of genotype-phenotype data. However, as we enter a new era of accelerating genomic technology, PharmGKB has expanded to accommodate large-scale genotype and phenotype data, such as gene and exon expression data (<affymetrix.com>). The integration of this data into PharmGKB is essential to provide a global view and understanding of genetic variants and their associated phenotypes. Just as we integrate external resources to help understand the functional role of variants or genotype data, it is also important to capture large-scale phenotype data (e.g., expression data), to better understand the global implications of each variant or combination of variants. PharmGKB is focused on aggregating data that are pharmacogenomically relevant,

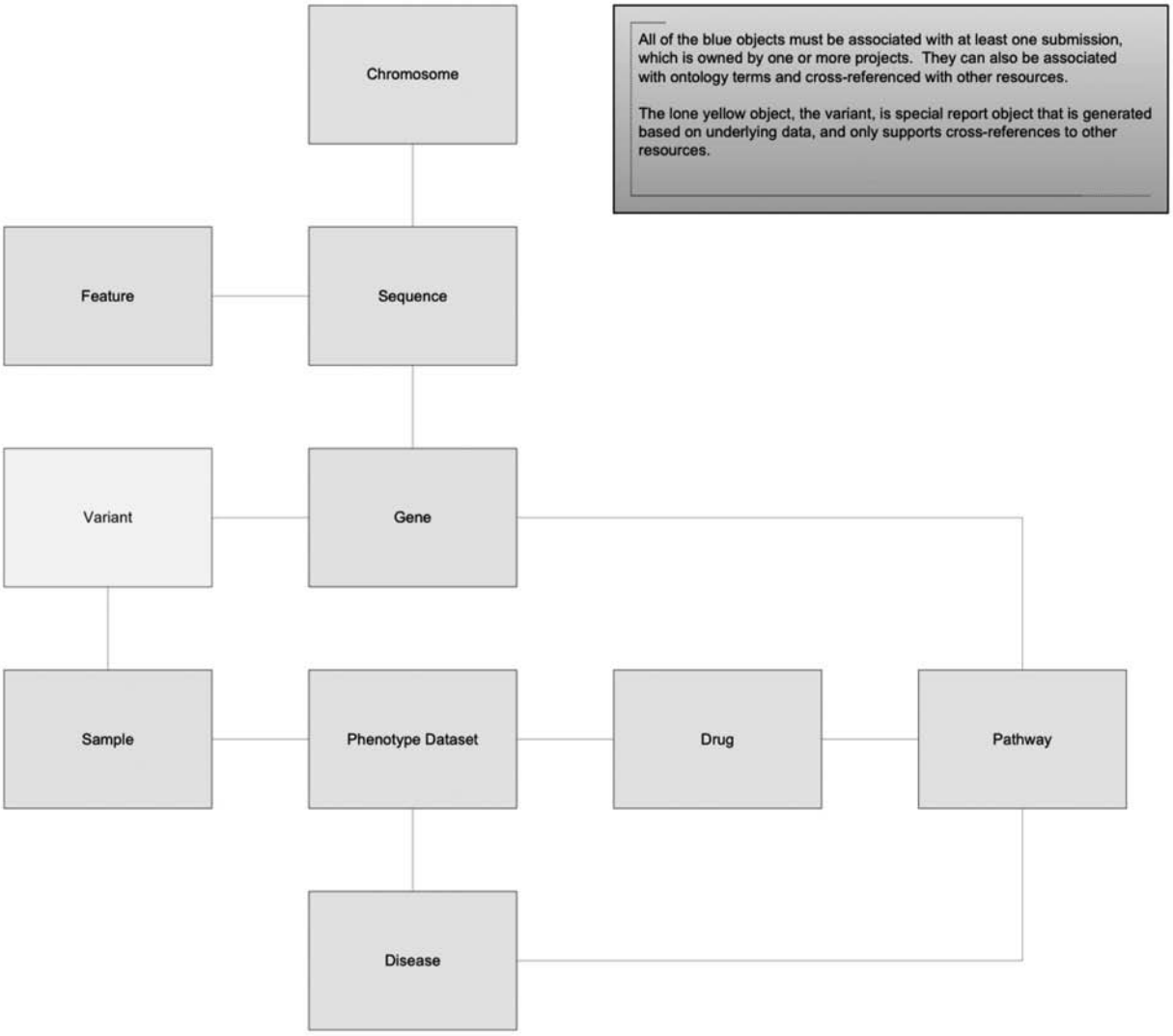
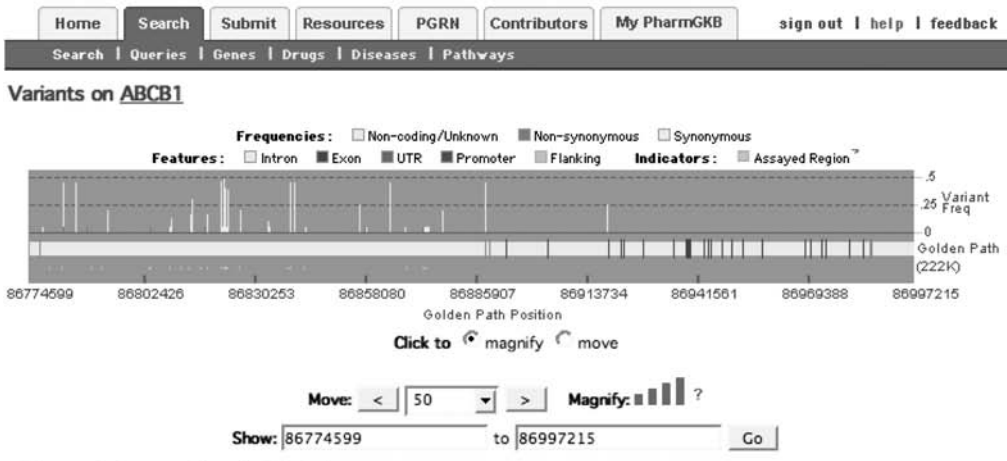


FIG. 2. An overview of PharmGKB's XML schema.

regardless of data type and format, which includes data reporting genotype-phenotype relationships, drug variability, and pharmacodynamic and pharmacokinetic datasets. PharmGKB is already capable of handling and disseminating large-scale genome-wide genotyping data, so the addition of the large-scale phenotypes is essential to complete the equation.

LARGE-SCALE GENOTYPE DATA

Advances in genomic technology have catalyzed research in population variation (Bonnen et al., 2006; Carlson, 2006). These technologies provide methods to manipulate and analyze genomic information in order to identify genes involved in disease processes. Polymorphic markers are identified by these technologies and may be used to locate genomic regions associated with disease phenotypes. In particular, SNPs are the most frequent type of variation in the human genome, occurring about once every several hundred base pairs. Currently there are over 11 million SNPs that have been reported to dbSNP, a public repository



GP Position	dbSNP Id	Variant	Feature	Amino Acid Translation	Frequency	Number of Chromosomes	Assay Types	Flags	Data
chr7:86778189	rs28364275	T/C	3' UTR		96.91%/3.09%	486	PCR		View
chr7:86778286		C/A	Exon	**Thr/Lys**	99.79%/0.21%	486	PCR		View
chr7:86778302	rs28364274	G/A	Exon	**Val/Ile**	99.79%/0.21%	486	PCR		View
chr7:86783296	rs1045642	C/T	Exon	-/-	53.86%/46.14%	1734	RFLP		View
chr7:86783310	rs17149694	T/A	Exon	**Ser/Thr**	95.31%/4.69%	490	PCR		View
chr7:86783409		T/C	Exon	**Trp/Arg**	99.8%/0.2%	490	PCR		View
chr7:86786465	rs6949448	A/G	Intron		55%/45%	20	SNP Array		View
chr7:86789329	rs28401798	C/G	Exon	**Pro/Ala**	99.8%/0.2%	494	PCR		View
chr7:86794573	rs2235067	G/A	Intron		80%/20%	20	SNP Array		View
chr7:86794712	rs28401776	G/A	Intron		99.8%/0.2%	494	PCR		View

FIG. 3. The PharmGKB Gene Browser and Variant Page for ABCB1. The Gene Browser shows the relative position of gene structures, the locations of variants including their frequencies and properties. The table of variants, below the Genome Browser contains detailed information about variants found at each GoldenPath position. Additional links allow one to drill down to flanking sequences and frequencies broken down by ethnicity and individual genotyped data.

for which the rarer SNP allele has a frequency of at least 1% (International-HapMap-Consortium, 2005; Sherry et al., 2001). These markers are abundant in the genome, highly reported and well studied, and are now amenable to analysis by high-throughput genotype assays or SNP arrays. SNP arrays provide an opportunity for studying genetic variation in genome-wide analyses (Cargill et al., 1999; Hernandez-Boussard et al., 2006).

These large-scale genotyping technologies have increased the amount of sequence variant data being generated by at least two to three orders of magnitude. As the prices decrease, these data are appearing increasingly in the peer-reviewed literature. Both Affymetrix and Illumina (<illumina.com>) have recently come out with technologies that can analyze 3,000 to 500,000 SNP markers in a single assay (<affymetrix.com>) (Gunderson et al., 2005). These platforms have been successful in several different types of pharmacogenomic studies, including linkage analysis studies, loss of heterozygosity studies, comparative genomic hybridization studies and genetic association studies (Bonnen et al., 2006; Craig and Stephan, 2005). These studies have made the largest impact on genetic association studies, allowing researchers to quickly and efficiently compare thousands to markers, which previously required extreme amounts of time and money. Genetic association studies are utilizing the larger SNP arrays (from 100,000 to 500,000 SNPs per array) and are emerging with a better understanding of the association of genetic variants with complex disease/phenotypes. These arrays have left researchers with very high expectations con-

cerning the ability to perform gene association studies to identify susceptibility genes for complex diseases using a comprehensive map of human genetic variation and allele association.

PHENOTYPE CLASSIFICATION

The relationship between genotype data and a phenotype is rarely straightforward and, most often, complex. This complexity arises because a single genetic variant does not usually produce a medically relevant phenotype, except in a very few monogenetic diseases. More often, one poorly defined phenotype is associated with one or more variants. Several genotypes might be sufficient to explain impaired function in the gene product, but still may not be adequate to explain all characteristics of a given phenotype. Therefore, the precise phenotype definition used in a genotype-phenotype association study must be clearly defined for accurate interpretation, aggregation, and replication. For human studies this is exceptionally challenging because there is not a single ontology that can adequately describe all categories of phenotypic evidence. To address this deficiency, PharmGKB has created four general categories of phenotypic measurements used in pharmacogenomic studies. These categories include molecular and functional assays, pharmacokinetics, pharmacodynamics and drug response, and clinical outcomes (Table 1). This categorization is not an ontology—it is a "tagging" system that aids in the integration and aggregation of diverse datasets. We are working toward a separate ontology for each code of evidence, such as SNOMED for clinical outcome and MedDRA for adverse drug reactions (Brown et al., 1999). There is still a need for a Hu-

Table 1. Categories of Pharmacogenetic Knowledge

CO: Clinical outcome	Genetic variations in the response to drugs can cause measurable differences in clinical endpoints, such as rates of cure, morbidity, side effects, and death. Data in this category demonstrate that genetic variability in the context of a drug effect significantly changes medical outcomes. These data sets are different from pharmacodynamics data sets, which may show a difference that is not sufficiently significant to alter practice or policy.
PD: Pharmacodynamics and drug response	Genetic variation in drug targets can cause measurable differences in the response of an organism to a drug. Data in this category document that the biological or physiologic response to a drug varies and that this variation can be associated with the variation of one or more genes. This variation often is measured at the whole-organism level. The measured variables may be surrogates for clinical responses, but do not constitute outcomes themselves.
PK: Pharmacokinetics	Genetic variation in processes involved in the absorption, distribution, metabolism, or elimination of a drug can result in changes in drug availability. Data in this category are primarily concerned with demonstrating that genetic polymorphisms lead to variations in the levels or concentrations of drugs or their metabolites at the site of action.
FA: Molecular and cellular functional assays	Genetic variation can alter results of molecular and cellular functional assays, which may correlate with variations in the organism's drug response. Data in this category demonstrate associations between genetic variation and laboratory assays of function at the molecular or cellular level. These assays may test the molecular properties of drug targets or drug-metabolizing enzymes, or may test the cellular properties of cells involved in the response to a drug (such as whole-cell gene expression).
GN: Genotype	Genotype is the internally coded, heritable information carried by the organism. Variation in genotype is fundamental to pharmacogenetics and is measured as sequence variation in individual genes—the type and location of the variation and the frequency of the variation in the populations of interest. This genetic variation is independent of individual drugs, but forms the basis for variation in response to drugs.

man Phenome Project that describes and catalogs all phenotypes attributed to sequence variation (Scriver, 2004). A large part of the human genome remains phenotypically uncharacterized and is likely to encode many unexpected and novel phenotypes (Brinkman et al., 2006).

Environmental influences, which are major confounding factors when determining genotype-phenotype relationships, are often excluded from the genotype-phenotype discussion. Environmental or nongenetic influences have been shown to play a key role in factors such as drug absorption, distribution, metabolism, and excretion. Perhaps these factors might best be addressed through metabonomic studies (Griffin, 2006; Need et al., 2005; Shastry, 2006). Regardless of the method of analysis, they should be included. Using metabonomic studies, variant drug responses could be predetermined and therefore avoided based on an individual's metabolic signatures (Clayton et al., 2006). These types of drug-related phenotypes are clearly relevant to pharmacogenomic research and should be considered when performing meta-analyses of genomic data, particularly drug variant responses.

EXPRESSION DATA

Expression data, both gene and exon arrays, provide a global view of sequence variation and gene activity and allow us to identify key differences between samples. These arrays have become an important analytical tool for genomic research, such as the successful identification of new markers that predict drug response (Ayers et al., 2004; Gyorffy et al., 2005, 2006). Expression arrays are also used as discovery tools to identify new oncogenes and tumor suppressor genes that can potentially identify new drug targets (Garraway et al., 2005; Solit et al., 2006). Exon arrays are now available to additionally address alternative splice sites on the expression of genes and give more insight into the underlying mechanisms involved in the differential expression of genes. These data are large-scale, with new exon arrays producing millions of data points per array. Microarray data are abundant in the scientific literature, and it is certain that exon array data will follow the same path. It is therefore crucial that these data be integrated into pharmacogenomic analyses.

INTEGRATION OF PHARMACOGENOMIC DATASETS

Given all the different data types that are relevant to pharmacogenomics, it is clear that complex genomic data should be integrated into one system in order to catalyze pharmacogenomic research. The scale, complexity, and information content of these datasets make integration challenging. However, in order to get a global view of the pharmacodynamic or pharmacokinetic reactions, these datasets must be brought together in a timely manner for global analyses.

There are several factors that contribute to the complexity of data integration; one major factor is the size of the data being produced. The new exon arrays have over 1.4 million probes on one array, and 1 array is approximately 250 MB (~160 MB for 1 .CEL file and ~90 MB for the annotation data). As most experiments contain multiple samples, there are serious concerns about the computer memory capacity required to normalize these arrays across samples. A single experiment with 100 samples can easily produce tens if not hundreds of gigabytes of data, so the normalization and analyses are extremely memory demanding. SNP arrays are also a challenge for scalability. The larger arrays now measure up to 500,000 SNPs in a single assay. A medium-sized high-throughput array is approximately 95 MB for one sample (12 MB for the image file and ~84 MB for the raw data intensity values). Gene expression datasets require approximately 3–5 MB of storage. One can see that for genome-wide aggregated analyses of these datasets, the size of the data is prohibitive, with extreme requirements on hardware, even as the price of hard drives has plummeted. The real challenge lies in finding ways to efficiently traverse the layers of data.

DATA AGGREGATION AND ANNOTATION

The application of high-throughput functional genomics technologies has assisted in the generation of new types of hypothesis and biological questions and has enabled many complex hypotheses to be tested. There is a need for a uniform way to annotate the data across these various genomic technologies. In order to retrieve these data for meta-analyses, the terms describing the experiments need to be organized in some type of ontology. To facilitate the merging of data across communities, PharmGKB is taking advantage of an on-going effort, The Functional Genomics Investigation Ontology (FuGO). This project is being developed as a collaborative, international effort to provide a common resource of terms to annotate functional genomics investigations, as well as assays that generate large amounts of data and metadata (Whetzel et al., 2006). The investigators involved in the FuGO project represent both biological and technological domains. The project will facilitate interoperability with other open biological Ontologies (OBI). Using an umbrella ontology for the metadata, such as FuGO, will aid in the search and retrieval of relevant data across domains.

Although the FuGO will aid in capturing metadata for functional genomic experiments, there still lacks a standardized form for reporting human phenotypes, certainly in the area of disease and clinical outcomes as previously mentioned. Very often with lower-throughput studies, there is no easy way to group the measured phenotypes. These raw phenotype measurements may need to be redefined or mixed and matched to achieve greater precision. However, without some type of standardization there are myriad ways to report a clinical outcome, such as blood pressure. Without standardization, none of the phenotypes can be aggregated or even effectively searched. In addition, there are many datasets that look at phenotypes and drug affects in model organisms, such as the Iconix Pharmaceuticals (<iconixpharm.com>) dataset that used microarrays to analyze chemogenomic information in relation to pharmacology and toxicology in rat (Fielden et al., 2005; Ganter et al., 2005). This information should be integrated with the human data, using appropriate ontologies and data standards. However, the complex relationship between a phenotype and a genotype adds yet another layer of intricacy when aggregating these data together to perform meta-analyses.

The real opportunity stemming from the use of ontologies for phenotype data is the possibility of summing up the data for whole genome annotation. It is important to capture why individual datasets are selected and what the relevant conclusions are from each individual dataset. For example, it has been shown that in prostate cancer patients, several genes are differentially expressed in some patients with poor diagnosis (Pollack et al., 2003). With the aggregated data (e.g., exon array and genotyping data), one would be able to perhaps pinpoint a specific exon of that gene that is differently expressed in a subset of patients, and, burrowing down further, a specific sequence polymorphism that is changing the function of the gene can yield valuable information on the causative effects of the differential expression. These types of functional genomic questions, while not assessable via a single study, can be answered by canned analyses at sites that have different levels of genome-wide data, such as PharmGKB. Having access to a variety of drug-related genomic datasets that are globally annotated will be invaluable for catalyzing new research.

CONCLUSIONS

The number of data available from genomic studies is rapidly growing. There is a new interest in associating genetic variants with drug responses or personalized medicine that demands the conglomeration of these genomic data for analysis as a whole. The integration, aggregation, and annotation of these data together in a timely manner is challenging, and the scalability of these data produce major challenges for a knowledge base, such as PharmGKB. However, it is only through the global meta-annotation of these data that we will understand the relationship between molecular changes in therapeutic targets and drug response in individuals. This understanding is crucial for establishing the clinical relevance of genomic approaches.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support by grants from NIGMS, NHLBI, NHGRI, NIEHS, NCI, and NLM within the NIH, HHS (U01GM61374) for the PharmGKB.

REFERENCES

AYERS, M., SYMMANS, W.F., STEC, J., DAMOKOSH, A.I., CLARK, E., HESS, K., et al. (2004). Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. J Clin Oncol **22,** 2284–2293.

BONNEN, P.E., PE'ER, I., PLENGE, R.M., SALIT, J., LOWE, J.K., SHAPERO, M.H., et al. (2006). Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. Nat Genet **38,** 214–217.

BRINKMAN, R.R., DUBE, M.P., ROULEAU, G.A., ORR, A.C., and SAMUELS, M.E. (2006). Human monogenic disorders—a source of novel drug targets. Nat Rev Genet **7,** 249–260.

BROWN, E.G., WOOD, L., and WOOD, S. (1999). The medical dictionary for regulatory activities (MedDRA). Drug Saf **20,** 109–117.

CARGILL, M., ALTSHULER, D., IRELAND, J., SKLAR, P., ARDLIE, K., PATIL, N., et al. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet **22,** 231–238.

CARLSON, C.S. (2006). Agnosticism and equity in genome-wide association studies. Nat Genet **38,** 605–606.

CLAYTON, T.A., LINDON, J.C., CLOAREC, O., ANTTI, H., CHARUEL, C., HANTON, G., et al. (2006). Pharmacometabonomic phenotyping and personalized drug treatment. Nature **440,** 1073–1077.

CRAIG, D.W., and STEPHAN, D.A. (2005). Applications of whole-genome high-density SNP genotyping. Expert Rev Mol Diag **5,** 159–170.

FIELDEN, M.R., PEARSON, C., BRENNAN, R., and KOLAJA, K.L. (2005). Preclinical drug safety analysis by chemogenomic profiling in the liver. Am J Pharmacogenomics **5,** 161–171.

GANTER, B., TUGENDREICH, S., PEARSON, C.I., AYANOGLU, E., BAUMHUETER, S., BOSTIAN, K.A., et al. (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. J Biotechnol **119,** 219–244.

GARRAWAY, L.A., WIDLUND, H.R., RUBIN, M.A., GETZ, G., BERGER, A. J., RAMASWAMY, S., et al. (2005). Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature **436,** 117–122.

GRIFFIN, J.L. (2006). The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? Philos Trans R Soc Lond B Biol Sci **361,** 147–161.

GUNDERSON, K.L., STEEMERS, F.J., LEE, G., MENDOZA, L.G., and CHEE, M.S. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet **37,** 549–554.

GYORFFY, B., SERRA, V., JURCHOTT, K., ABDUL-GHANI, R., GARBER, M., STEIN, U., et al. (2005). Prediction of doxorubicin sensitivity in breast tumors based on gene expression profiles of drug-resistant cell lines correlates with patient survival. Oncogene **24,** 7542–7551.

GYORFFY, B., SUROWIAK, P., KIESSLICH, O., DENKERT, C., SCHAFER, R., DIETEL, M., et al. (2006). Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. Int J Cancer **118,** 1699–1712.

HERNANDEZ-BOUSSARD, T., KLEIN, T.E., and ALTMAN, R.B. (2006). Pharmacogenomics: the relevance of emerging genotyping technologies. MLO Med Lab Obs **38,** 24, 26–30.

HIRAKAWA, M., TANAKA, T., HASHIMOTO, Y., KURODA, M., TAKAGI, T., and NAKAMURA, Y. (2002). JSNP: a database of common gene variations in the Japanese population. Nucleic Acids Res **30,** 158–162.

INTERNATIONAL-HAPMAP-CONSORTIUM (2005). A haplotype map of the human genome. Nature **437,** 1299–1320.

KENT, W.J., SUGNET, C.W., FUREY, T.S., ROSKIN, K.M., PRINGLE, T.H., ZAHLER, A.M., et al. (2002). The human genome browser at UCSC. Genome Res **12,** 996–1006.

KLEIN, T.E., and ALTMAN, R.B. (2004). PharmGKB: the pharmacogenetics and pharmacogenomics knowledge base. Pharmacogenomics J **4,** 1.

NEED, A.C., MOTULSKY, A.G., and GOLDSTEIN, D.B. (2005). Priorities and standards in pharmacogenetic research. Nat Genet **37,** 671–681.

POLLACK, A., COWEN, D., TRONCOSO, P., ZAGARS, G.K., VON ESCHENBACH, A.C., MEISTRICH, M.L., et al. (2003). Molecular markers of outcome after radiotherapy in patients with prostate carcinoma: Ki-67, bcl-2, bax, and bcl-x. Cancer **97,** 1630–1638.

SCRIVER, C.R. (2004). After the genome—the phenome? J Inherit Metab Dis **27,** 305–317.
SHASTRY, B.S. (2006). Pharmacogenetics and the concept of individualized medicine. Pharmacogenomics J **6,** 16–21.
SHERRY, S.T., WARD, M.H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E.M., et al. (2001). dbSNP: the NCBI database of genetic variation. Nucl Acids Res **29,** 308–311.
SOLIT, D.B., GARRAWAY, L.A., PRATILAS, C.A., SAWAI, A., GETZ, G., BASSO, A., et al. (2006). BRAF mutation predicts sensitivity to MEK inhibition. Nature **439,** 358–362.
TATE, S.K., and GOLDSTEIN, D.B. (2004). Will tomorrow's medicines work for everyone? Nat Genet **36,** S34–42.
WHETZEL, P.L., RYAN R. BRINKMAN, R.R., CAUSTON, H.C., FAN, L., FOSTEL, J., FRAGOSO, G., et al., and members of the communities collaborating in the Fugo Project (2006). The development of FuGO—an ontology for functional genomics investigations. OMICS **10,** 199–204.

Address reprint requests to:
Russ B. Altman
Stanford University Medical School
Department of Genetics
300 Pasteur Drive, L301
Stanford, CA 94305-5120

E-mail: russ.altman@stanford.edu