Published in final edited form as:

Pharmacogenet Genomics. 2010 April ; 20(4): 269–273. doi:10.1097/FPC.0b013e328337b8d6.

***PACdb*: a database for cell-based pharmacogenomics**

Eric R. Gamazon^a, Shiwei Duan^{a,f}, Wei Zhang^a, Rong Stephanie Huang^a, Emily O. Kistner^{a,e}, Mary Eileen Dolan^{a,c,d}, and Nancy J. Cox^{a,b}

^aDepartment of Medicine, The University of Chicago, Chicago, Illinois

^bDepartment of Human Genetics, The University of Chicago, Chicago, Illinois

^cCommittee on Clinical Pharmacology and Pharmacogenomics, The University of Chicago, Chicago, Illinois

^dCancer Research Center, The University of Chicago, Chicago, Illinois

^eDepartment of Medicine, Division of Biostatistics and Epidemiology, Medical University of South Carolina, Charleston, South Carolina, USA

^fSingapore Institute for Clinical Sciences, A-STAR, Singapore

Abstract

We have developed Pharmacogenomics And Cell database (*PACdb*), a results database that makes available relationships between single nucleotide polymorphisms, gene expression, and cellular sensitivity to various drugs in cell-based models to help determine genetic variants associated with drug response. The current version also supports summary analysis on differentially expressed genes between the HapMap samples of European and African ancestry, as well as queries for summary information of correlations between gene expression and pharmacological phenotypes. At present, data generated on the following anticancer agents are included: carboplatin, cisplatin, etoposide, daunorubicin, and cytarabine (Ara-C). The database is also available to assist in the investigation of the effects of potential confounding variables (e.g. cell proliferation rate) in lymphoblastoid cell lines. *PACdb* will be regularly updated to include more drugs and new datasets (e.g. baseline microRNA levels). *PACdb* will be linked into PharmGKB to benefit the next wave of pharmacogenetic and pharmacogenomic discovery.

Keywords

drug response; gene expression; HapMap; lymphoblastoid cell line; pharmacogenomics; single nucleotide polymorphism

Introduction

Lymphoblastoid cell lines (LCLs) derived from targeted populations have been used to evaluate genetic factors responsible for a variety of clinical phenotypes. The International

© 2010 Wolters Kluwer Health | Lippincott Williams & Wilkins

Correspondence to Mary Eileen Dolan, PhD, 900 E. 57th St., KCBD Rm 7100, The University of Chicago, Chicago, IL 60637, USA, Tel: +1 773 702 4441; fax: +1 773 702 9268; edolan@medicine.bsd.uchicago.edu .

Eric R. Gamazon, Shiwei Duan and Wei Zhang contributed equally to this study

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website (www.pharmacogeneticsandgenomics.com).

Conflicts of interest: none declared.

HapMap Project [1], by releasing a human haplotype map on a panel of Epstein–Barr virus-transformed LCLs from major continental populations (CEU: Caucasians from Utah, USA; YRI: Yoruba people from Ibadan, Nigeria; CHB: Han Chinese from Beijing, China; JPT: Japanese from Tokyo, Japan), has opened up new avenues for understanding the relationship between genetic variation, particularly in the form of single nucleotide polymorphisms (SNPs), and complex traits or phenotypes including gene expression and drug response. Particularly, with the availability of extensive genotypic data as well as the potential for LCLs to be a good model system for hematological toxicities and possibly other toxicities, the HapMap samples provide a means to interrogate the genetic causes of interindividual and interethnic differences in drug-induced cytotoxicity [2–7].

Notably, variation in drug response phenotype is likely to be multifactorial, involving the interaction of various genetic and nongenetic factors. Separating the genetic contribution from confounding factors (e.g. concomitant medications) poses considerable challenges. In oncology, the situation is compounded by difficulties, both ethical and practical, of conducting clinical trials with drugs that cause serious toxicities on unaffected family members for genetic studies. Cell-based pharmacogenomics aims to address some of the challenges associated with uncovering the genetic basis of interindividual variation in drug response. Although the cell-based model comes with its own limitations (e.g. tissue specificity), there is a need for a resource that both uses the benefits of cell-based pharmacogenomics and assists in the investigation of confounding factors.

In addition to HapMap's catalyzing role, advances in genotyping technologies, expression microarrays, and bioinformatic tools have begun to allow unbiased, genome-wide approaches for dissecting the genetic architecture of gene expression, which resides between DNA sequence variation and clinical or cellular phenotypes (e.g. drug response). Studies using LCLs suggest that gene expression is a quantitative trait accounted for by *cis*-acting and *trans*-acting common genetic variants, and that common genetic variants contribute to differential expression between populations of individuals from different geographic locations [8].

However, the unprecedented volume of genome-wide data – genotype (e.g. >3.1 million HapMap SNPs [1]), gene expression (e.g. >13 000 transcript clusters in the Affymetrix exon array data [9]), and phenotype (e.g. drug response) – presents new challenges, particularly in computational resources required to analyze, interpret, and prioritize results related to the identification and characterization of genetic variants that predict for a pharmacological outcome. Pharmacogenomics And Cell database (*PACdb*) was developed to address some of these unmet computational and interpretational challenges. Investigators involved in pharmacogenomic research may benefit from *PACdb* as a discovery platform or as a validation dataset for clinical observations.

Database description

Technical architecture—*PACdb* (<http://www.PACdb.org/>) is written in PHP in a three-tiered architecture intended to provide versatility, scalability, and flexibility. This approach segments the infrastructure into three distinct logical layers – front-end, middle-tier, and data storage – such that each layer can be maintained and developed independently. The front-end is a web client application that accepts user input and query selection. The middle-tier is a collection of objects – drug, SNP, gene, expression, phenotype, population, among others – responsible for the execution of queries. The data are stored in a relational database using MySQL. Figure 1 shows the data served by the current *PACdb* and its potential extensibility.

Genotype–cytotoxicity association—Currently, *PACdb* serves data on the following drugs: carboplatin [10], cisplatin [3], etoposide [4,5], daunorubicin [2], and cytarabine (Ara-C) [6]. *PACdb* will be updated regularly to include other drugs tested using this model. Genome-wide association studies of genotype and cytotoxicity were performed for each of the drugs. SNP genotypes were retrieved from the HapMap database Version 23a (<http://www.hapmap.org/>) for the CEU and YRI samples. SNPs with Mendelian allele transmission errors in autosomes in the 30 CEU and 30 YRI trios were excluded, resulting in approximately 2 million common SNPs with minor allele frequency greater than 5% each for CEU and YRI. Cytotoxicity was measured using cell growth inhibition after treatment with increasing concentrations of drug. The concentration necessary to inhibit 50% of cell growth (IC₅₀), the area under the cellular survival curve (AUC) and percent cellular growth inhibition at specified concentrations were the phenotypes analyzed. Quantitative Transmission Disequilibrium Test [11] was performed separately on each quantitative phenotype (i.e. IC₅₀ etc.) within each population, to detect genotype–cytotoxicity associations. Multiple comparisons were adjusted for by QVALUE [12]. The Genotype–Cytotoxicity Query tool allows the user to specify a *P* value (uncorrected) and/or *q* value threshold and retrieves a list of SNPs that show associations with drug-induced cytotoxicity for a particular drug in a particular population (Fig. 2). *PACdb* highlights a SNP annotation system that uses associations with cellular phenotypes (including test statistics, uncorrected *P* values and false discovery rates) as an important approach to characterize pharmaco-SNP functionality, which supplements publicly available annotation systems often solely based on genomic location and population-level parameters (e.g. minor allele frequency).

Genotype–gene expression association, population differential gene expression, and alternative splicing—The initial dataset includes gene expression data on 176 HapMap cell lines (87 CEU and 89 YRI) evaluated on the Affymetrix exon array [9]. Quantitative Transmission Disequilibrium Test analysis was performed, separately within each population as before, to assess the relationship between approximately 2 million common SNPs and gene expression (> 13 000 expressed transcript clusters) in LCLs. *PACdb* highlights the importance, in pharmacogenomic discovery, of gene expression as a mechanism of drug response variation. In conjunction with our SCAN database [13] (<http://www.SCANDb.org/>), which holds HapMap SNP associations to transcriptional expression, *PACdb* may be used to interrogate whether SNPs significantly associated with a drug's cytotoxicity are also regulators of certain gene expression phenotypes. Earlier, differential expression of transcript clusters between the CEU and YRI samples was also evaluated [9]. The Differential Expression Query tool (Supplemental Fig. S1, <http://links.lww.com/FPC/A128>) allows the user to retrieve these differentially expressed genes based on specified cutoffs.

Alternative splicing, a posttranscriptional mechanism that generates proteomic diversity, is likely to contribute to drug response variation. The Splicing Index Query Tool (Supplemental Fig. S2, <http://links.lww.com/FPC/A129>) provides, for any gene, a list of splicing index values, which are probeset-level expression intensities normalized by transcript-level expression intensities, so that these normalized intensities may be compared across sample groups:

$$SI_{ij} = p_{ij} / t_i$$

where p_{ij} is the normalized intensity of the j th probeset (exon level) of the i th transcript cluster and t_i is the normalized intensity of the i th transcript cluster (gene level) [14]. In the absence of splicing, exon-level expression should 'match' gene-level expression. High or low splicing index (e.g. based on an arbitrary cutoff of >1.2 or <0.8) for a probeset enables

the detection of splice events, though further evaluations may be necessary for selecting a more reliable cutoff. The tool thus provides the ability to retrieve the set of potential splice events for a list of queried genes.

Gene expression–cytotoxicity association—For each drug and each population, a general linear model using a Toeplitz covariance structure with two diagonal bands to allow for familial dependencies in the data was constructed between $\log_2$ -transformed gene expression level [9] and $\log_2$ -transformed cytotoxicity measures (i.e. AUC or IC_{50}) as the dependent variable. In this model, each trio is a unit; mother and father IC_{50} are treated as independent while the offspring IC_{50} is allowed to covary with both father and mother. In case of independent samples, this model reduces to the familiar linear regression. QVALUE [12] was also used to control false discovery rate. The Expression-Phenotype Query Tool allows the user to select a drug, enter a list of transcript clusters or a list of genes, as well as a P value (uncorrected) and/or q value cutoff. The result set shows association between drug phenotype and gene expression, the uncorrected P value, as well as the relevant transcript cluster (Supplemental Fig. S3, <http://links.lww.com/FPC/A130>).

Genotype–quantitative phenotype association—A general whole-genome approach using cell-based models can be used to elucidate genetic variants contributing to a wide range of cellular phenotypes (e.g. apoptosis). *PACdb* has been designed to be extensible to other phenotypes relevant for cell-based pharmacogenomics and to other analytic approaches (Supplemental Fig. S4, <http://links.lww.com/FPC/A131>). The candidate-gene approach focuses the search for associated genetic variants on a particular gene with a priori knowledge of its function. In contrast, whole-genome approaches (e.g. linkage analysis) use an unbiased procedure for the discovery of functional genetic variants. A combination of the two approaches may, in practice, be necessary in pharmacogenetic studies. *PACdb*, as a platform for pharmacogenomic discovery and validation, was designed to incorporate diverse analytic approaches.

An example of application

As an example of the potential utility of *PACdb*, we demonstrated that *PACdb* can be used to find some previously reported associations in pharmacogenetics. Particularly, it has been shown, in cancer cell lines and patients with cancers, acquired resistance to cisplatin (e.g. in ovarian carcinomas) can be mediated by secondary mutations in the tumor suppressor *BRCA2* [15]. A *PACdb* query (P value cutoff = 0.05) using the Gene Expression–Cytotoxicity Association Tool shows that *BRCA2* expression significantly predicts cisplatin IC_{50} ($P = 0.021$, $q = 0.23$); furthermore, a follow-up query using the Genotype–Cytotoxicity Association Tool (P value cutoff = 0.01, population = CEU) generated multiple SNPs within *BRCA2* whose genotypes showed significant (e.g. rs206077 at $P = 9 \times 10^{-3}$, $q = 0.71$) associations with both cisplatin IC_{50} and AUC. Therefore, *PACdb*, as a comprehensive approach for showing relationships between genotype, drug response, and gene expression, may also serve as a platform for further validations or confirmations of earlier findings.

Discussion

PACdb is designed to allow the incorporation of multiple phenotypes (e.g. drug response, cytotoxicity, apoptosis, or enzyme activity), and/or analytic approaches in cell-based pharmacogenomic studies, particularly those using the LCL models. *PACdb* implements a comprehensive frame-work for pharmacogenomic discovery and validation that integrates studies of the transcriptome, pharmacologic phenotypes (e.g. IC_{50}), and genetic variation (e.g. SNPs). It allows the user to ‘modulate’ the result set, according to the level of

significance desired or the proportion of false positives [12] when a particular test is called significant.

Some examples of *PACdb* usage include:

1. Querying SNPs contributing to drug-induced cytotoxicity through effects on gene expression (see Ref. [4] for example of the general approaches used to identify pharmacogenomic loci using LCLs);
2. Querying SNPs important in the cytotoxicity of one drug to see whether those SNPs are important in cellular sensitivity to any other drugs;
3. Querying SNPs identified in clinical studies associated with response to chemotherapy for their association with cellular sensitivity to these chemotherapeutic agents.

Although the current version of *PACdb* serves results based on HapMap SNP genotypic data and five anticancer drugs, its design and implementation facilitate the incorporation of pharmacogenomic datasets (e.g. baseline microRNA levels, copy number variation data, epigenetic data, and more drugs) from other genome-wide association studies as well as other analytic approaches to characterize genetic variants. The availability of more detailed maps of human genetic variation (e.g. the 1000 Genomes Project and the SeattleSNPs Project data on LCLs) will benefit the next wave of pharmacogenomic studies, therefore these datasets could potentially be incorporated in *PACdb* in the future. Finally, *PACdb* will be linked into existing databases, particularly the Pharmacogenomics Knowledge Base (PharmGKB, <http://www.PharmGKB.org/>), to allow the pharmacogenetic and pharmacogenomic community to leverage the available pharmacology-related resources.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported through the Pharmacogenetics of Anticancer Agents Research Group (<http://www.pharmacogenetics.org/>) by the NIH/NIGMS grant U01GM61393 with data deposits supported by U01GM 61374 (<http://www.pharmgkb.org/>), NIH/NCI Breast SPORE P50 CA125183, NIH/NCI grants CA136765 and R21CA139278; University of Chicago Cancer Research Center Pilot Funding.

References

1. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–861. [PubMed: 17943122]
2. Huang RS, Duan S, Kistner EO, Bleibel WK, Delaney SM, Fackenthal DL, et al. Genetic variants contributing to daunorubicin-induced cytotoxicity. *Cancer Res* 2008;68:3161–3168. [PubMed: 18451141]
3. Huang RS, Duan S, Shukla SJ, Kistner EO, Clark TA, Chen TX, et al. Identification of genetic variants contributing to Cisplatin-induced cytotoxicity by use of a genomewide approach. *Am J Hum Genet* 2007;81:427–437. [PubMed: 17701890]
4. Huang RS, Duan S, Bleibel WK, Kistner EO, Zhang W, Clark TA, et al. A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc Natl Acad Sci U S A* 2007;104:9758–9763. [PubMed: 17537913]
5. Bleibel WK, Duan S, Huang RS, Kistner EO, Shukla SJ, Wu X, et al. Identification of genomic regions contributing to etoposide-induced cytotoxicity. *Hum Genet* 2009;125:173–180. [PubMed: 19089452]

6. Hartford CM, Duan S, Delaney SM, Mi S, Kistner EO, Lamba JK, et al. Population-specific genetic variants important in susceptibility to cytarabine/arabinoside cytotoxicity. *Blood* 2009;113:2145–2153. [PubMed: 19109566]
7. Huang RS, Kistner EO, Bleibel WK, Shukla SJ, Dolan ME. Effect of population and gender on chemotherapeutic agent-induced cytotoxicity. *Mol Cancer Ther* 2007;6:31–36. [PubMed: 17237264]
8. Cheung VG, Spielman RS. Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nat Rev Genet* 2009;10:595–604. [PubMed: 19636342]
9. Zhang W, Duan S, Kistner EO, Bleibel WK, Huang RS, Clark TA, et al. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet* 2008;82:631–640. [PubMed: 18313023]
10. Huang RS, Duan S, Kistner EO, Hartford CM, Dolan ME. Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans. *Mol Cancer Ther* 2008;7:3038–3046. [PubMed: 18765826]
11. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000;66:279–292. [PubMed: 10631157]
12. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100:9440–9445. [PubMed: 12883005]
13. Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, et al. SCAN: SNP and Copy Number Annotation. *Bioinformatics* 2010;26:259–262. [PubMed: 19933162]
14. Duan S, Bleibel WK, Wisel SA, Huang RS, Wu X, He L, et al. Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum Genet* 2009;125:81–93. [PubMed: 19052777]
15. Sakai W, Swisher EM, Karlan BY, Agarwal MK, Higgins J, Friedman C, et al. Secondary mutations as a mechanism of cisplatin resistance in *BRCA2*-mediated cancers. *Nature* 2008;451:1116–1120. [PubMed: 18264087]

Fig. 1. Pharmacogenomics And Cell database (*PACdb*) serves results from cell-based pharmacogenomic studies. CEU, Caucasian residents from Utah, USA; YRI, Yoruba people from Ibadan, Nigeria. Dashed arrows indicate future developments (e.g. more drugs and datasets).

The screenshot shows the PACdb website interface. At the top is the PACdb logo. Below it is a navigation bar with links: Home, Search, Resources, SCAN, and Contact Us. On the left side, there is a vertical menu with the following options: Cytotoxicity / Genotype Association (highlighted), Differential Expression, Gene Expression / Phenotype Association, and Splicing Index. The main content area is titled 'Cytotoxicity / Genotype Association'. It contains several input fields: 'Select drug:' with a dropdown arrow, 'Select population:' with a dropdown menu showing 'All', and 'Enter SNPs:' with a large text area. Below these, there is an 'or' separator, followed by 'Upload SNP list:' with a text input field and a 'Browse...' button. Further down, there are two more input fields: 'Enter p-value threshold:' with the value '0.0001' entered, and 'Enter q-value threshold:' which is empty. At the bottom of the form are two buttons: 'Submit' and 'Reset'.

Fig. 2.

Genotype–Cytotoxicity Association Query Tool. With drug as input, single nucleotide polymorphisms (SNPs) that show association at a user-specified threshold are output, along with the relevant phenotype (e.g. area under the cellular survival curve and IC_{50}). Each SNP is annotated with host gene, function, and genomic RefSeq position. The results, as true of all Pharmacogenomics And Cell database (*PACdb*) queries, may be exported to a spreadsheet for further analysis.