APPLICATION NOTE **Genomics & Informatics** Vol. 7(3) 163-167, September 2009

The Korean Pharmacogenomic Database at NIFDS: 2008 Update

T.S. Kang[1], S.W. Woo[2], H.J. Park[1], S.Y. Han[1], M.H. Park[1] and M.W. Chung[1]*

[1]Clinical Research Division, Toxicological Evaluation and Research Department, Korea Food and Drug Administration, Seoul 122-704, Korea, [2]Pharmaceutical Management Division, Korea Food and Drug Administration, Seoul 122-704, Korea

Abstract

Since its first release in 2007, the National Institute of Food and Drug Safety Evaluation (NIFDS) has provided pharmacogenomic and comparative information specific to Koreans to allow regulatory reviewers and researchers to adapt their working practices to pharmacogenomics. The highlights of this year's additions include "Drug Information", "Gene Information" and "Pharmacogenomic information in the drug labels" sections. These new additions provide information on 737 genes, 719 drugs and pharmacogenomic data of the labels or relabels of 253 approved drugs as of November 2008. The latest version of the Korean Pharmacogenomic Database (KPD, release 2.0) has expanded significantly since its previous release. More SNP and haplotype information has been added to the database with the latest version of the KPD containing approximately four times as many SNPs and haplotypes than the previous version (719 vs. 152, and 30 vs. 7 respectively). Through the "SNP" and "Haplotype" sections, the KPD provides unique Korean SNP and haplotype information as well as comparative information of other populations (Japanese, Chinese, European, African) to offer a range of pharmacogenomic data that can help reviewers and the public understand pharmacogenomic information. The quality and quantity of information in the KPD has also been improved considerably. This data can be found at: http://www.nitr.go.kr/nitr/contents/m134700/view.do/.

Keywords: database, Korean people, pharmacogenomics

*Corresponding author: E-mail mwchung@kfda.go.kr
Tel +82-2-380-1807, Fax +82-2-388-6458
Accepted 24 July 2009

Introduction

The Internet is an ideal platform for storing, providing and exchanging large-scale genetic information, and there are a large number of web-based bioinformatics sites worldwide. These databases are operated by many organizations for a variety of purposes, and offer huge volumes of information online for public access (Chang Kug Kim 2009; Hongseok Tae 2009; Innocenti 2005; Werner Kalow 2002).

The most well-known pharmacogenomic information database is the Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB, www.pharmGKB.org), which is the most comprehensive pharmacogenomic information database in the world. The PharmGKB provides information on genotypes, molecular and functional assays, pharmacodynamics, pharmacokinetics, and clinical outcomes to those involved in pharmacogenomic research (Innocenti 2005).

However, this database contains a wide range of information, and beginners can find it difficult to obtain the required data from this information labyrinth. First-time visiting researchers find it extremely difficult to locate the desired information on the PharmGKB. There are no databases that provide unique Korean information on the SNP and Haplotypes in drug-metabolism-related genes. In addition, there is no information on the relationships between particular SNPs and the drug metabolism, which can be used to identify candidate genes affecting disease progression or drug reactions.

The Korean Pharmacogenomic database (KPD) was developed to allow easier searches for Korean Pharmacogenomic information (Tae Sun Kang *et al*., 2008), and is located at the National Institute of Food and Drug Safety Evaluation (NIFDS) homepage. When the KPD was first released in 2007, it offered the selected and specialized data (mainly SNP and haplotype information in liver metabolizing enzymes, trasporters and receptors). For the convenience of the public, researchers and scientific reviewer, link functions to major intermational databases (NCBI, dbSNP database, and PharmGKB) were established in the KPD.

Since its first release, the KPD has been used in a wide range of applications including drug evaluations, selection of SNPs in Koreans and general pharmacogenomic education. The KPD is being updated con-

tinuously to provide more practical pharmacogenomic information for drug evaluations. This paper reports these developments as well as the many additions and improvements appearing in the latest version of the KPD (release 2.0)

Methods

Database Organization

Oracle 10g was used for database management system and the interface between database and web was implemented on JAVA/JSP. Details related to the genotype information, linkage disequilibrium blocks, haplotype structures and database organization on the KPD have been described previously (Tae Sun Kang *et al.*, 2008).

Drug and label information

Drug and pharmacogenomic information in drug label were obtained from the Physician's Desk Reference (2007) for the U.S.A, the National Health Service homepage (http://emc.medicines.org.uk/) for the U.K, Drugs in Japan (2007) for Japan and DIMS and EZDRUG for Korea respectively.

Quality Assurance

For completeness of the database, each data used in the KPD is entered or prepared by one menber of the clinical research division in NIFDS and separately validated by second member of the clinical research division. Additional spot checks are routinely performed on each entry by senior members of the review group, including two phamacists, a statistician and a PhD-level genetist.

Results

The KPD was developed to collect genetic information unique to the Korean population, and to allow easy information searches within the numerous databases available. It is also expected to act as a bridge that facilitates easier information gathering through links to larger databases that provide more precise information. With the increasing need to provide more practical information to evaluate drugs as well as educate scientific reviewers and the public, three new data field additions have been added and database size and coverage has been expanded.

Data field additions

As shown in Table 1, the KPD has three new additions.

Gene Information section

The "Gene information section" was established to provide information on genes, such as those encoding the major drug-metabolizing enzymes, transporters, and receptors, which can affect the drug response (adverse effects and efficacy), based on the SNPs provided in the "SNP information" section. A total of 737 genes are arranged in alphabetical order to allow easy access, and the number of relevant genes is recorded next to each entry. This section also offers detailed information on the summary, key pathways, drug and substrates, important variants and important haplotypes of each gene (Fig. 1). For those who require more specialized information on each gene, quality information is provided by links to the PharmGKB database (www.pharmGKB.org), the UCSC Genome Bioinformatics (http://genome.ucsc.edu) and the OMIM at NCBI (http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim) sites.

Drug Information section

The "Drug Information" section was established to provide drug information, based on the drugs provided in the "Pharmacogenomic Information in the drug labels" section. A total of 592 drugs have been arranged in alphabetical order to allow easy access, and the number of relevant drugs is recorded next to each entry. This section also offers detailed information on the molecular weight, indications, mechanisms of action, absorption data, distribution data, protein binding data, biotransformation data, half life, elimination data, adverse effects and interactions of each drug (Fig. 1). Those requiring more specialized information on each drug can obtain it via links to the PharmGKB database (www.pharmGKB.org/) site.

Pharmacogenomic Information in the drug labels section

The "Pharmacogenomic Information in the drug labels" section was established to provide information on the

Table 1. Comparison of the data content in KPD release 1.0 with that in KPD release 2.0

Category	Release 1.0	Release 2.0
Gene Information	0 gene	737 genes
Drug Information	0 drug	592 drugs
Pharmacogenomic Information in the drug labels	0 drug	253 drugs
SNP Information	152 genes	719 genes
Haplotype Information	4 genes	10 genes

Fig. 1. Screenshot of some of the new additions to the KPD including Gene information, Drug information and Pharmacogenomic information in drug label.

use of pharmacogenomic data in drug labels and relabels in Korea and foreign countries. A total of 253 approved drugs have been arranged in alphabetical order to allow easy access, and the number of relevant drugs is recorded next to each entry. This section allows detailed use of the pharmacogenomic data listed in the Physicians' Desk Reference. Through Valid Biomarker (V.B.) information, pharmacogenomic data in the drug labels and relabels in Korea can be compared with those of other regulations (the U.S.A, the U.K and Japan) (Fig. 1).

Expanded database size and coverage

Table 1 shows a detailed comparison of KPD release 1.0 with KPD release 2.0. The latest release of the KPD has detailed information on the SNPs of a total of 719 genes. Comparisons of the SNP frequencies in the Korean population with those of other races registered in the International HapMap database (Europeans, Chinese, Japanese, and African) are also provided.

The "haplotype information" section has been also updated to provide information on 10 unique Korean haplotypes. Korean linkage disequilibrium block data for the genes encoding the major metabolic-enzymes is included. The blocks comprise the SNPs with minor allele frequency (MAF) values $>5\%$ using the Gabriel method in the Haploview 3.32 program (Ardlie, Kruglyak and Seielstad 2002; Barrett, Fry, Maller and Daly 2005). The haplotype frequencies of the major liver metabolic-enzymes for the Korean population can be compared with those of other races registered in the International HapMap (Europeans, Chinese, Japanese, and African). In addition, the latest version contains approximately four times as much pharmacogenomic information as the previous release (30 versus 7) through the "Pharmacogenomic Information and Education" facility within the KPD.

Discussion

NIFDS has been operating the KPD with the main aim of offering users a "bridging" function to simplify the search for useful information. The KPD has systematically collected and maintained unique Korean pharmacogenomic information from literatures and other databases and provided scientic reviewers and researchers with the selected data. Since its first release, the KPD has provided data on Korean single-nucleotide polymorphisms (SNPs) and haplotypes. It also allows a comparison of the Korean SNP and haplotype frequencies with those of the other ethnic groups registered in the International HapMap (Thorisson, Smith, Krishnan and Stein 2005). The KPD is focused primarily on providing the detailed pharmacogenomic data needed to facilitate drug evaluation. Therefore, in the latest version, three new additions ("Gene Information" section, "Drug Information" section and "Pharmacogenomic information in the drug labels" section) were established, and the data coverage and database linkages were expanded to provide more practical pharmacogenomic information for drug evaluations. These new additions provide scientific reviewers and the general public with the concise information needed to understand ethnic differences and evaluate bridging studies. It is also expected to be useful for finding unique Korean Pharmacogenomic data without the need to search other databases. The KPD contains extensive links to almost all major bioinformatics and pharmacogenomic databases (the dbSNP and OMIM at NCBI, PharmGKB and UCSC sites) (Karolchik *et al*., 2008; Kuhn *et al*., 2007; Kuhn *et al*., 2009; Wheeler *et al*., 2008). Through these links, the KPD provides convenient search functions to obtain more precise information with just a single mouse click. What is the most important thing is to update the database with the laststt pharcogenomic data on a constant basis in order to maintain the KPD effectively. Local expert organizations and networks are needed to guarantee the reliability of the pharmacogenomic information contained in the database. In order to ensure that the KPD is used effectively by scientific reviewers and the general public, the types of visitors to the site must be assessed and the database must be upgraded periodically to reflect the demands of the major visiting population. Overall, it is expected that the KPD will serve as a useful resource to both scientific reviewers and the general public.

Acknowledgements

The authors wish to thank the ISTECH INC. and Dr. Yang-Seok, Kim who contributed to development of the KPD. This study was supported by grant no. 08171-KFDA504 from the Korea Food and Drug Administration.

References

Ardlie, K.G., Kruglyak, L., and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet*. 3, 299-309.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263-265.

Innocenti, F. (2005). The pharmacogenetics and pharmacogenomics knowledge base. HUMANA Press.

Karolchik, D., Kuhn, R.M., Baertsch, R., Barber, G.P., Clawson, H., Diekhans, M., Giardine, B., Harte, R.A.,

Hinrichs, A.S., Hsu, F., Kober, K.M., Miller, W., Pedersen, J.S., Pohl, A., Raney, B.J., Rhead, B., Rosenbloom, K.R., Smith, K.E., Stanke, M., Thakkapallayil, A., Trumbower, H., Wang, T., Zweig, A.S., Haussler, D., and Kent, W.J. (2008). The UCSC genome browser database: 2008 update. *Nucl. Acids Res*. 36, D773-779.

Kim, C.K., Jeon, Y.A., Cho, G.T., Kwon, S.W., Kim, Y.H., and Hong, S.B. (2009). An intergrated database and web service for microbial resources at KACC. *Genomics Inform*. 7, 41-45.

Kuhn, R.M., Karolchik, D., Zweig, A.S., Trumbower, H., Thomas, D.J., Thakkapallayil, A., Sugnet, C.W., Stanke, M., Smith, K.E., Siepel, A., Rosenbloom, K.R., Rhead, B., Raney, B.J., Pohl, A., Pedersen, J.S., Hsu, F., Hinrichs, A.S., Harte, R.A., Diekhans, M., Clawson, H., Bejerano, G., Barber, G.P., Baertsch, R., Haussler, D., and Kent, W.J. (2007). The UCSC genome browser database: update 2007. *Nucl. Acids Res*. 35, D668-673.

Kuhn, R.M., Karolchik, D., Zweig, A.S., Wang, T., Smith, K.E., Rosenbloom, K.R., Rhead, B., Raney, B.J., Pohl, A., Pheasant, M., Meyer, L., Hsu, F., Hinrichs, A.S., Harte, R.A., Giardine, B., Fujita, P., Diekhans, M., Dreszer, T., Clawson, H., Barber, G.P., Haussler, D., and Kent, W.J. (2009). The UCSC genome browser database: update 2009. *Nucl. Acids Res*. 37, D755-761.

Tae Sun Kang, S.W.W., Hyun Joo Park, and Jaesook Roh. (2008). Korean pharmacogenomic database: pharmacogenomic haplotype analysis of the Korean population and ethnic comparisons. *Biomol. Ther*. 15, 77-81.

Tae, H., Han, J.M., Ahn, B.Y., and Park, K. (2009). Biostore: a repository system for registering and distributing public biology databases. *Genomics Inform*. 7, 49- 51.

Thorisson, G.A., Smith, A.V., Krishnan, L., and Stein, L.D. (2005). The international HapMap project web site. *Genome Res*. 15, 1592-1593.

Werner Kalow, U.A.M.a.R.F.T. (2002). WWW Bioinformatics Resources. Taylor & Francis.

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Miller, V., Ostell, J., Pruitt, K.D., Schuler, G.D., Shumway, M., Sequeira, E., Sherry, S.T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R.L., Tatusova, T.A., Wagner, L., and Yaschenko, E. (2008). Database resources of the national center for biotechnology information. *Nucl. Acids Res*. 36, D13-21.