

# Korean Pharmacogenomic Database: Pharmacogenomic Haplotype Analysis of the Korean Population and Ethnic Comparisons

Tae Sun KANG, Sun Wook Woo, Hyun Joo PARK, and Jaesook ROH<sup>1\*</sup>

Department of Pharmacological Research, National Institute of Toxicological Research, Seoul, Korea

<sup>1</sup>Department of Obstetrics & Gynecology, College of Medicine, Hanyang University, Seoul, Korea

(Received April 12, 2008; Accepted June 17, 2008)

---

**Abstract** – The major aim of the Korean Pharmacogenomic Database (KPD) is to offer to users a “bridging” function, making the search for useful information easier. This database has also been established to collect unique Korean genotype data from other databases and to directly link these data to other major databases that offer more informative data. In this way, searches for information about new drug developments and easier and faster evaluation of the more complex and larger databases are possible. The KPD is located at the National Institute of Toxicological Research homepage (<http://www.nitr.go.kr/nitr/contents/m134700/view.do>), and offers Korean single-nucleotide polymorphism (SNP) information for 154 genes and haplotype information. It also compares the Korean SNP and haplotype frequencies with those of the other ethnic groups registered in the International HapMap. Through the Pharmacogenomic Information and Education facility, we also provide evaluators and the public with information about the concept of pharmacogenomic information, research trends, and the drug regulations of other countries. Because the drug responses of Koreans are not necessarily the same as those of Chinese or Japanese people, it is expected that the systematic operation of the KPD will allow the definition of racial differences and various genomic biomarkers (haplotypes or SNPs) for use in bridging studies and in the approval of new drugs.

**Keywords:** Korean people, database, pharmacogenomics

---

## INTRODUCTION

All around the world, numerous bioinformatics sites are operated for different purposes. For example, there are diverse types of sites, run by organizations such as universities, government organizations, and general companies, which support related sites, interest areas (proteins, nucleotide sequences, diseases, etc.), and groups with various functions (sequence research, structure database searches, analysis, and education) (Tang and Helmeeste, 2002). Among these diverse bioinformatics sites, databases that provide information on single-nucleotide polymorphism (SNP) markers, which constitute over 90% of all human genetic variation, are gradually increasing in importance (Lee and Han, 2006). Because of the importance of SNPs, an international consortium, the International HapMap Project, was established to identify SNPs and examine their characteristics. It also provides infor-

mation on linkage disequilibrium (LD) blocks and haplotype maps of the human genome. The haplotype blocks and common haplotypes that occur in Africans, Europeans, and Asians (Japanese/Chinese) have been examined and a haplotype map (genome-wide HapMap) that includes the genomes of all humans has been constructed (Thorisson *et al.*, 2005).

Moreover, in 1999, the National Institutes of Health (NIH) recognized the need for researchers to have free access to accepted genotypic and phenotypic data from pharmacogenetic and pharmacogenomic studies, and thus established the Pharmacogenetics Research Network (PGRN). As the major resource of the PGRN, a knowledge-based database called the Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB, [www.pharmGKB.org](http://www.pharmGKB.org)) was formed, which provides diverse information on genotypes, molecular and functional assays, pharmacodynamics, pharmacokinetics, and clinical outcomes to those involved in pharmacogenomic research (Caroline *et al.*, 2005).

Because of the importance of SNP markers, diverse

---

\*Corresponding author

Tel: +82-2-2290-8404, Fax: +82-2-2296-8472

E-mail: rohjaesook@hanyang.ac.kr

studies have also been conducted in Korea, but the connections between each SNP database are tenuous, and most databases focus primarily on the association between SNP genotyping data and disease. In particular, there is a lack of databases that provide information on the haplotypes associated with specific diseases and on drug-metabolism-related genes, or gene frequency comparisons between races. Nor is there information on the relationships between particular SNPs and drug metabolism, which can be used to identify candidate genes affecting disease progression or drug reactions.

In this study, we aimed to draw together the haplotype-based SNP information unique to the Korean population and to systematically provide tag SNP information. Furthermore, by comparing the SNP information of particular drug-metabolism-related genes and the haplotype frequencies of different races, ethnic difference data can be used for drug development and evaluation.

## MATERIALS AND METHODS

### Genotype information

The genotype data for Japanese (45 people), Chinese (45 people), European (90 people), and African (90 people) populations, registered on the International HapMap ([www.hapmap.org](http://www.hapmap.org)), were used as the representative genotype information for other races. The genotype information for the Korean population (N=50~100) was obtained from the Knowledge Base for Korean Pharmacogenomics Research Network (<http://kprn.snubi.org>), operated by Seoul National University.

### Analysis of LD blocks and haplotype structures

All the Hardy–Weinberg  $P$  values of the genotype markers used in the LD and haplotype analyses were less than the HapMap standard value of 0.0010. The LD block and haplotype structures were analyzed using the Haploview 3.32 program (Barrett *et al.*, 2005). In the LD analysis, the confidence interval method was used (with the basic algorithm of a 95% confidence interval using the  $D$  Prime value) and “informative comparison” values exceeding 95% were defined as having “strong LD” (Barrett *et al.*, 2005). The haplotype structure analysis was defined using the “accelerated expectation-maximization (EM) algorithm” (Ardlie *et al.*, 2002). The tag SNP within the LD block was defined with the Tagger function (using the aggressive tagger method) in the Haploview 3.32 program.

### Database Organization

Oracle 10 g was used for database management sys-

tem and the interface between database and web was implemented on JAVA/JSP.

## RESULTS

The purpose of the KPD in terms of drug development and licensing is to collect genetic information unique to the Korean people, and to allow easier information searches within the numerous available databases, which are gradually becoming larger and more complex. It is also expected to act as a bridge that facilitates easier information gathering through links to larger databases that provide more precise information.

The KPD is located at the National Institute of Toxicological Research (NITR) website (<http://www.nitr.go.kr/nitr/contents/m134700/view.do>). It provides unique Korean SNP frequency information for 154 types of genes, as well as haplotype analysis and frequency information based on the SNP database. Various educational materials relating to pharmacogenomics are also provided.

### SNP information section

Within the KPD, the “SNP information section” has been newly established to provide unique Korean SNP information relating to genes such as those encoding the major drug-metabolizing enzymes, transporters, and receptors, which can affect drug reactions (adverse effects and efficacy). These SNPs do not include information about association with disease and adverse drug reactions, but are selected in the order of high frequency (minor allele frequency above 5%) in Korean people. A total of 154 genes have been arranged in alphabetical order to allow easy access, and the numbers of relevant genes are recorded next to each entry. It also offers detailed information on the SNP frequencies (or allele frequencies) in the Korean population, chromosomal positional information (POS) for the relevant SNPs, and the unique registration number (rs NO) of each SNP. Comparisons of the SNP frequencies in the Korean population and those of other races registered in the International HapMap database (Europeans, Chinese, Japanese, and African) are also provided. For those who require more specialized information regarding each SNP, quality information is provided through links to the dbSNP database (<http://www.ncbi.nlm.nih.gov>) and the UCSC Genome Bioinformatics (<http://genome.ucsc.edu>) sites (Fig. 1).

### Haplotype information section

Within the KPD, the “Haplotype information section”

약물유전정보 DB Home / 정보마당 / 독성과학원 DB 검색 / 약물유전정보 DB

페이지프린트

Gene Drug **SNP정보란** Haplotype 정보란 PDR 약물유전정보 교육자료실

**CYP2C9** 목록

cytochrome P450, family 2, subfamily C, polypeptide 9

Alternate Names	P450 MP-4; P450 PB-1; P450IIC9; cytochrome P-450 S-mephenytoin 4-hydroxylase; cytochrome P450, subfamily IIC (mephenytoin 4-hydroxylase), polypeptide 10; cytochrome P450, subfamily IIC (mephenytoin 4-hydroxylase), polypeptide 9; flavoprotein-linked monoo
Alternate Symbols	1; CPC9; CYP2C; CYP2C10; CYP2C9*1; CYP2C9*1A; MGC88320; P450 MP-4; P450 PB-1; P450IIC9

SNP 정보 (한국인 및 타인종과의 빈도수 비교자료)

rNO	POS	feature	한국인			유럽인			중국인			일본인			아프리카인		
			allele	frequency	N	allele	frequency	N	allele	frequency	N	allele	frequency	N	allele	frequency	N
rs7900194	chr10:96692056	exon	A/G	100/0	24												
rs9332130	chr10:96699027	exon	A/G	100/0	24	A/G	99,1/0,9	116	A/G	100/0	88	A/G	100/0	86	A/G	100/0	116
	chr10:96731043	Exon	A/C	98,2/1,8	389												
rs9332239	chr10:96738767	exon	C/T	100/0	24	C/T	100/0	116	C/T	100/0	88	C/T	100/0	86	C/T	100/0	116
rs9332092	chr10:96686519	flanking	C/T	95,83/4,17	24	C/T	5,8/94,2	120	C/T	4,5/95,5	88	C/T	3,4/96,6	88	C/T	0/100	120

**Fig. 1.** Screenshot of “SNP information” showing SNP frequencies in the Korean and other populations registered in the International HapMap.

has been newly established to provide information on unique Korean haplotypes, based on the SNPs provided in the “SNP information section.” These data are also arranged in alphabetical order of the relevant genes for easier searches, and the numbers of relevant genes are recorded next to each entry.

**LD Analysis:** In the section “LD analysis information”, the uniquely Korean LD block data are included for the major liver metabolic-enzymes. The blocks comprise SNPs with minor allele frequency (MAF) values above 5%, using the Gabriel method in the Haploview 3.32 program (Ardlie *et al.*, 2002; Barrett *et al.*, 2005).

**Haplotype and tag:** In the section “Haplotype and tag information”, information is provided on how the haplotypes in the Korean population are combined and how their frequencies are calculated based on the SNPs in the LD blocks.

**Haplotype comparisons with other races and frequency information:** The haplotype frequencies of the major liver metabolic-enzymes for the Korean population have been compared with those of other races. The haplotypes are constructed from the SNP genotypes of the same genes in the other races registered in the International HapMap (Europeans, Chinese, Japanese, and African), and the haplotype frequencies of each race are

compared with those of the Korean population.

### Operation of the Pharmacogenomic Information and Education facility

Pharmacogenetic information is provided to evaluators and the general public through the “Pharmacogenomic Information and Education” facility within the KPD. Within this facility, the five submenus include “the pharmacogenomic information section”, “the need for pharmacogenetic information”, “related terms”, “related sites”, and “educational materials”, which facilitate easy access to and utilization of the pharmacogenetic data by the general public and evaluators (Fig. 2).

### DISCUSSION

The Internet is an ideal platform on which to store, provide, and exchange large-scale genetic information, and there are numerous web-based bioinformatics sites worldwide. These databases are operated by different organizations for different purposes, and offer huge quantities of information online for public access (Caroline *et al.*, 2005; Tang and Hemeste, 2002).

Because the current online databases contain a very wide range of information, beginners find it difficult to

The screenshot shows the NITR website interface. At the top, there is a search bar with the text '통합검색' and '검색어를 입력하세요.' Below it, a navigation menu includes '정보마당', '자료실', '보도정책자료', '학습정보', and '독성과학원 DB 검색'. The main content area is titled '약물유전정보 DB' and includes a search bar with '제목' and '검색' buttons. Below the search bar, there is a table with the following data:

번호	제목	담당부서	등록일	첨부	조회수
10	단일염기다형성 다중분석법 - nanogene	임상약리팀	2008-01-25		3

**Fig. 2.** Screenshot of “Pharmacogenomic Information and Education”, which allows easy access to, and utilization of, pharmacogenomic information by the general public and evaluators.

gather the required data from this information labyrinth. The most renowned bioinformatics database is that of the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>, hereinafter NCBI), which is the most comprehensive database in the world, established in 1988. The NCBI site links numerous types of bioinformatics databases: gene banks, molecular databases, genomic biology resources, and software resources. Large quantities of information are found not only at the NCBI site, but also in the various linked databases that are available to researchers and the general public worldwide (Wheeler *et al.*, 2005).

For first-time visiting researchers, it is extremely difficult to find the desired information in the complex sea of information on the NCBI database. For the general public especially, it is virtually impossible to know where to start searching when visiting the site, and how to deal with the extensive data.

There is a growing need for a database that can act as a bridge to facilitate easier searches for information. It is becoming increasingly important to facilitate simpler information searches within this sea of information through links to larger databases that provide more precise information. The KPD at the NITR website has moved away from the complex database characteristics described

above, and provides only selective and specialized data (mainly SNP information relating to liver metabolic enzymes, transporters, and receptors, and haplotype information), offering easier information search functions to the general public and evaluators. Through links with major international databases (NCBI, dbSNP database, International HapMap database, and PharmGKB), it provides convenient search functions for more precise information with just one click.

Furthermore, the Pharmacogenomic Information and Education facility within the KPD offers educational data to those encountering pharmacogenetic information for the first time and to the general public, who may find searching databases difficult because of their lack of background knowledge. Diverse pharmacogenetic information is provided for drug designers to utilize in their drug evaluations with increased convenience, because of the links to other relevant sites.

For this pharmacogenomic information database to be used most effectively, it is necessary to identify the latest domestic and overseas data trends and to undertake continuous management of the database. Furthermore, to guarantee the reliability of the pharmacogenomic information within the database, local expert organizations and networks are required. To ensure that the KPD is

used actively by evaluators, as well as by the general public, the types of visitors to the site must be assessed and the database must be periodically upgraded to reflect the demands of the major visiting population.

### ACKNOWLEDGMENTS

We wish to thank Yoonsook Lee for the general management of this project including valuable conception of the ethnic comparison and the ISTECH INC. and Dr. Yang-Seok Kim for their contribution in IT support to development of KPD. This work was supported by grant no. 07151KFDA674 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. *Nature Reviews Genetics*. **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.
- Caroline, F. T., Teri, E. K. and Russ, B. A. (2005). The Pharmacogenetics and Pharmacogenomics Knowledge Base. In *Pharmacogenomics Methods and Protocols* (Federico Innocenti, Ed.), pp. 179-191. HUMANA Press.
- Lee, J. K. and Han, H. L. (2006). Linkage disequilibrium and HapMap. In *Genetic Variation and Diseases* (J. K. Lee, Ed.), pp. 70-89. Worldscience.
- Tang, S. W. and Helmeste, D. M. (2002). WWW Bioinformatics Resources. In *Pharmacogenomics* (Werner Kalow, Urs A. Meyer and Rachel F. Tyndale, Ed.), pp. 493-514. Taylor & Francis.
- Thorisson, G. A., Smith, A. V., Krishnan, L. and Stein, L.D. (2005). The International HapMap Project Web site. *Genome Res*. **15**, 1592-1593.
- 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., Helmsberg, 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. (2005). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. **33**, 39-45.