Published in final edited form as:

Curr Protoc Bioinformatics. 2008 September ; 0 14: Unit14.7. doi:10.1002/0471250953.bi1407s23.

PharmGKB, an Integrated Resource of Pharmacogenomic Data and Knowledge

Li Gong¹, Ryan P. Owen¹, Winston Gor¹, Russ B. Altman^{1,2}, and Teri E. Klein¹

¹Genetics Department, Stanford University, Stanford, California

²Department of Bioengineering, Stanford University, Stanford, California

Abstract

The PharmGKB (<http://www.pharmgkb.org>) is a publicly available online resource that aims to facilitate understanding on how genetic variation contributes to variation in drug response. It is not only a repository of pharmacogenomics primary data, but also provides fully curated knowledge including drug pathways, annotated pharmacogene summaries and relationships amongst genes, drugs and diseases. This unit describes how to navigate the PharmGKB website to retrieve detailed information on genes and important variants, as well as their relationship to drugs and diseases. It also includes protocols on our drug-centered pathway, annotated pharmacogene summaries and our web services for downloading the underlying data. Workflow on how to use PharmGKB to facilitate design of the pharmacogenomic study is also described in this unit.

Keywords

Database; pharmacogenomics; pharmacogenetics; drug response; genetic variation; pathway analysis; SNP; polymorphisms; study design

INTRODUCTION

Pharmacogenomics is the study of how genetic variation contributes to variation in drug response. Driven by technology advancements in the post-genomic era, pharmacogenomics research has the potential to optimize drug efficacy and minimize toxicity. It bridges the gap between the scientific discoveries and clinical application, and offers the exciting promise of personalized drug therapy. The Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB) is a publicly available internet resource for pharmacogenomic data and knowledge (Klein and Altman, 2004). PharmGKB strives to capture rapid advancements in the pharmacogenomics area. It is the central data repository for pharmacogenetic and pharmacogenomic data, in addition to providing integrated knowledge including drug pathways, gene summaries, and relationships amongst genes, drugs, and diseases.

PharmGKB serves diverse user groups from the scientific community. It provides comprehensive and integrated drug, gene, and disease information to pharmacologists, clinical investigators, and biologists, as well as to informaticians. The PharmGKB homepage has been designed in a way that highlights the primary interests of most users, and registered users have complete access to individualized genotype and phenotype data for discovery research and further analysis. PharmGKB is also an excellent educational portal

for any person who is new to pharmacogenomics. A graphic schema depicting the central elements involved in drug response and associated genetic basis is displayed on the PharmGKB homepage. Also provided are lecture materials, tutorials, and useful links intended to help people familiarize themselves with the fundamental concepts of pharmacogenomics research and personalized medicine.

The protocols in this unit describe how to use PharmGKB to browse pharmacogenomic data and knowledge. The basic protocol describes how to navigate the PharmGKB homepage and browse through the knowledge base starting from search by gene. Support Protocol 1 explains in detail how to use our variant browser and variant table; Support Protocol 2 describes how to explore our unique drug-centered pathways; Support Protocol 3 demonstrates the types of knowledge contained in our Very Important Pharmacogene (VIP) gene summaries; and, Support Protocol 4 describes our web services project which allows our user to bulk download data from PharmGKB.

BASIC PROTOCOL: NAVIGATING THE HOMEPAGE OF PHARMGKB USING SEARCH BY GENE

This protocol will introduce the basic techniques used for searching and browsing the content on the PharmGKB Web site.

Necessary Resources

Hardware—Computer with an internet connection

Software—Any up-to-date Web browser will work.

Files—No input files required

The PharmGKB home page: Getting started

- 1 Open the PharmGKB homepage at <http://www.pharmgkb.org> in a web browser.

The PharmGKB homepage is the common entry point for all users. It has been designed to highlight the types of information that are most sought after by diverse groups of users. A graphic schema for understanding the basis of pharmacogenomics has also been included (Fig. 14.5.1).

The menu tabs at the top of the page provide access to the top-level section of the PharmGKB site; see Table 14.5.1 for a description of the various tabs. Prominently displayed in the center of the home page are clickable icons that allow our users to go directly to a specific type of pharmacogenomic-related data such as pathways, genes, variants of interest, drugs, diseases and download information. Right below the icon is the search box where a user can enter text for a Google-type query. The search box is also prominently displayed at the top right hand corner of the homepage. At the top right of the every PharmGKB

page, a frequently asked questions (FAQ) link and feedback link are available. A scientific curator responds to all feedback within 48 hr.

The PharmGKB home page also provides basic tutorial information about pharmacogenetics and pharmacogenomics. Below the search box is a graphic schema that illustrates the basic flow of pharmacogenetic information: After a drug is administered, it is absorbed, distributed, metabolized and excreted (pharmacokinetics; PK); the drug then reaches its target and elicits a drug response (pharmacodynamic effects; PD). Both the PK and PD of the drug can be influenced by an individual's genetic makeup (GN) and in turn, lead to distinct clinical outcomes (CO). The five categories of evidence (COE) mentioned above appear in this diagram, and their relationship to each other is indicated in the picture. Definitions of these terms are provided in the "Useful Links" section at bottom right of the homepage. Another valuable learning resource on the PharmGKB homepage is the list of Curators' Favorite Papers. This is a biweekly feature that covers the recent "hot" papers in pharmacogenomics. Each paper is annotated with the pertinent categories of evidence (COE) and tagged with relevant genes, drugs and diseases.

Searching by Gene and its associated variant, pathways, drugs and disease information

A gene can be searched by either typing the gene name or symbol in the search box, or, clicking on the gene icon from the home page and then browsing through the alphabetically sorted gene list. For example, to search for VKORC1, a key protein in vitamin K metabolism and target of the anticoagulant drug warfarin, type VKORC1 in the search box, then click enter. If no result is returned, try a synonym or partial name. Both alternative names and other symbols that might have been used in the literature for the gene of interest are included for all genes in PharmGKB. We adhere to the nomenclature at HUGO Gene Nomenclature Committee (HGNC) for official gene names (Eyre et al., 2006), and make every effort to keep them current.

2 Open the VKORC1 gene page (Fig. 14.5.2).

The main gene page is also organized by a tab system similar to the homepage. The overview tab lists the alternative gene symbols and gene names, as well as details such as gene and mRNA boundaries, and their OMIM phenotype if available. Additional tabs for the gene include Datasets, Pathways, Curated and non-curated publications. The last tab is the downloads/cross-references illustrates the links to download genotype or phenotype data associated with the gene. It also lists unique identifiers used by PharmGKB and other external genomic databases for the specific gene.

3 Click on the **VIP tab** to view the VKORC1 VIP gene summary for detailed information on variant and haplotype mapping and their importance in drug responses. (See support protocol 3 for details.)

- 4 Click on the **Variants** *tab* to display all the variants for VKORC1 available in PharmGKB in the browser, as well as in the variant table with variant details and functional annotation for variants of interests (see Support Protocol 1 for details).
- 5 To view curated phenotype data associated with VKORC1, click on the **Datasets** *tab*, then select the link titled **WUSTL warfarin dosing data** (Fig. 14.5.3).

Phenotype data at PharmGKB are organized by a tab system, similar to that on the homepage. The Overview tab lists the investigator, related genes, drugs, and disease, as well as a summary for the study; the second tab, Publications, lists all publications related to that phenotype; the third tab lists all column headers and descriptions for the individual phenotype data such gender, race, age, dose etc.; the Individualized data tab allows the user to view individualized subject data after the user logs in.

- 6 Click on the **Pathways** *tab* to view all pathways associated with VKORC1. Click on **Warfarin Pathway (PD)** to view the simplified diagram of the target of warfarin action and downstream genes and effects..

See Support Protocol 2 for details.

- 7 Click on the **Curated Publications** *tab* to see the manually curated literature information to find drugs and diseases associated with VKORC1. Click on **View** to see the **evidence of relationship** between drug (warfarin) and gene (VKORC1).
- 8 Click on **warfarin** in the **related drugs from the literature section** to bring up the drug page for warfarin, where the detailed pharmacology, mechanism of action, and therapeutic use of the drug are listed in the detail section of the page. Both the drug page and disease page follow a design similar to the gene page.
- 9 Click on **Atrial Fibrillation** in the **Related Diseases from Literature** section to see the disease page for Atrial Fibrillation.
- 10 To download genotype and phenotype data related to VKORC1, click on **Downloads/cross-references** *tab* on the gene page

All individualized primary data at PharmGKB are available for download by registered users. For bulk download of some or all of the data in PharmGKB for further analysis, please use our SOAP-based Web services (See support protocol 4 for details).

- 11 Under the Downloads/cross-references tab, click on links to go to external databases where addition information on VKORC1 gene may be found.

PharmGKB has established bidirectional links with leading gene, protein and drug resources, such as NCBI Entrez Gene 9; Maglott, et al. 2007), GeneCards (Safran et al., 2002), UniProtKB (Wu et al., 2006), and DrugBank (Wishart et al., 2006). We also provide links

from the gene page to *Online Mendelian Inheritance in Man (OMIM)*, Hamosh, et al., 2005), the *Genome Data Base (GDB)*, Letovsky et al., 1998), *NCBI RefSeq sequences* (Pruitt et al., 2007), and their associated *Gene Ontology* annotations (Harris et al., 2004). In the **Common Searches** section immediately below the *Cross-references*, users can check to see if their gene of interest is part of any pathway documented within public pathway databases such as the *Kyoto Encyclopedia of Genes and Genomes (KEGG)* (Kanehisa et al., 2004), *BioCarta* (<http://www.biocarta.com>), or *Reactome* (UNIT 8.7; Joshi-Tope et al., 2005).

SUPPORT PROTOCOL 1: ORIENTATION TO THE PHARMGKB VARIANT PAGE

The PharmGKB gene variant page contains a variant browser and a variant table. The variant browser displays all the polymorphisms in the gene of interest documented within various sources, such as PharmGKB primary data, single-nucleotide polymorphism (SNP) arrays (www.illumina.com and www.affymetrix.com), NCBI Single Nucleotide Polymorphism database (dbSNP, Sherry et al., 2001), and Japanese Single Nucleotide Polymorphisms Database (jSNP, Hirakawa et al., 2002). The variant table below the browser lists detailed non-array genotype data in PharmGKB such as their genomic positions, functional annotation for variants of interests, structural view of the coding variants, polymorphism frequencies, and assay types. Both the variant table and PharmGKB SNP array data for the gene of interest are available for download at the bottom of the variant page.

Necessary Resources

Hardware—Computer with an internet connection

Software—Any up-to-date browser will work

Files—No input files required

- 1 From the homepage (<http://www.pharmgkb.org>), click on the **genotyped genes** icon in the browse section. This will lead to the page for all genes with genotype information listed in alphabetical order.
- 2 Click on the letter V to go to all genes starting with letter V. The number “4” to the right of the letter V indicates the number of genes with variant data starting with that specific letter.
- 3 Click on the [variants] link to the right of VKORC1 to go to the variant gene page for VKORC1.
- 4 The variant gene page contains a **variant browser** at the top and a **variant table** below (Fig. 14.5.4).

The Variant browser gives a graphical representation of gene structure and location of variants contained within PharmGKB (including those derived from whole genome SNP arrays). Variants collected from external SNP databases such as dbSNP and jSNP are also available through the variant browser, allowing users to easily compare and contrast SNPs from different resources and identify regions that have a high density of polymorphisms. Each tick on the browser represents a variant from the respective resource. The gene features are also color coded to differentiate exons, introns, promoters, and untranslated regions (UTRs). By using the magnification and move tools below the browser, the user can move or zoom into the specific region of interest for a gene.

- 5 Scroll down the variant page to locate the **variant table** below the browser to see PharmGKB non-array variants and their genomic positions, functional annotations, frequencies, and assay types.

Clicking on the link in the “GP Position” column will open the UCSC Genome Browser (UNIT 1.4; Kuhn et al., 2007) in another window. Clicking on the link in the “dbSNP Id” column will open the dbSNP entry that corresponds to the variant in another window (UNIT 1.3). The entries in the “Feature” and “Amino Acid Translation” columns are derived from the default reference sequence from NCBI for the specific gene.

- 6 Click on the “G/A” variant (at GP position Chr16:31009822) in the variant column. This will display the variant report that includes the reference sequence for the specific variant.
- 7 Click on the “stars” under the variants of interest curation level to see the brief functional summary for the variants and their literature evidence support.

Stars are used to indicate the level of annotation applied to the variants: 1 star for non-curated annotations, 2 stars for curated annotations, and 3 stars for in-depth annotations. Note that non-curated variant information is accumulated solely by computational methods and has not been verified by the scientific staff at PharmGKB.

- 8 Click on the “Expanded Variants View” button below the variant browser to see the full variant table containing more in-depth information collected for the variants such as frequencies and assay types. Click on the link in the “Frequency” column (e.g., “58.97%/41.03%” at GP Position Chr16:31009822) to see a breakdown of the frequency across all variants reported by racial categories.

The value in the “Frequency” column is calculated by aggregating all variants reported at that Golden Path position. Clicking on the value will allow the user to drill down further for frequency data by race or ethnicity. The entry in the “Number of Chromosomes” column is

typically twice the number of subjects that were in the submitted sample set. Each variant is also annotated with the detailed genotyping assay performed on that variant in the “Assay type” column. The Phi link (Φ) in the “Flags” column indicates that phenotype data was collected on subjects genotyped for that Golden Path position.

- 8 Click on the [view](#) link under column “Data” in the full variant table to see subject genotype data reported at the specific position. (Individual level genotype data requires user registration.)
- 9 Scroll down to the bottom of the variant page to export the variant table in CVS/Excel/XML formats, along with any SNP array data from PharmGKB, for the gene of interest.

SUPPORT PROTOCOL 2: ORIENTATION TO THE PHARMGKB PATHWAY PAGES

The interactive drug-centered pathways displayed in PharmGKB provide an overview of how genes are involved in the pharmacokinetics (PK) and pharmacodynamics (PD) of drugs. The pathway diagrams use standard shapes and colors to represent genes, metabolites, drugs, and interactions. All genes and drugs on the pathway diagram are clickable. If the user clicks on these objects, the PharmGKB gene or drug page opens in a new browser window. Below the pathway picture is a “Description” of the pathway that describes the complex gene-drug relationships depicted in the pathway diagram. The pathway authors and the date of the most recent update are listed below the text of the description on the bottom of the pathway diagram. There is a section of useful links and downloads to the right of the pathway picture. At the top right, there is a link to return to the list of pathways. If PharmGKB has both PK and PD pathways for a given drug, the user will see a box with a drop down menu, allowing them to toggle between the PK and PD pathways. Also listed are links to related drugs, genes, and pathways that have been selected by the authors as being of potential interest to the user. Finally, on the lower right of the pathway page, there are download links for the evidence spreadsheet and the pathway image. The evidence spreadsheet describes each interaction depicted on the pathway, and it includes at least one peer-reviewed article with their PubMed reference identifier in support of each interaction.

Necessary Resources

Hardware—Computer with an internet connection

Software—Any up-to-date browser will work

Files—No input files required

1. From the homepage (<http://www.pharmgkb.org>), click on the pathway icon to access the list of pathways on PharmGKB. Pathways can also be accessed by clicking on the Search tab, then on Pathways.
2. Click on [Irinotecan Pathway](#) in the second page of the pathway lists to go to Irinotecan pathway (Fig. 14.5.5).

Irinotecan pathway shows the pharmacokinetic (PK) process of the chemotherapy drug irinotecan. Genes involved in biotransformation and transport of irinotecan are highlighted in this PK pathway. A pathway that describes the pharmacodynamic (PD) aspect of irinotecan is also available on PharmGKB titled “Irinotecan Pathway (Cancer)”. How irinotecan acts on its target topoisomerase I (TOP1) in the cancer cell is illustrated in this PD pathway.

3. Click on [Legend](#) on the top left corner of the pathway diagram to display the standard shapes that we use to designate the different objects in the pathway.
4. Click on the ABCC1 oval to go to the ABCC1 gene page.
5. Click on “irinotecan” will open a pop-up window of the drug page for irinotecan. Click on metabolite “SN38” to open a pop-up window which shows the chemical conversion from irinotecan to SN38, it also includes a link to the original article describing the conversion as well as a link back to the irinotecan drug page.
6. Click on the golden arrow between two objects (SN38 → SN38G) to see a link to the primary data titled “Irinotecan Clinical Data” that support the relationship. This pop up window also provides a link to the original article describing the influence of UGT1A1 genotype on the rate of glucuronidation of SN38 (PMID:12464801).
7. Click on the pull-down menu in the upper right hand corner of the pathway page to select alternative view of the pathway (liver or cancer for Irinotecan pathway, PK or PD in many other pathways).
8. Click on [Illustrator file](#) to download the pathway image in pdf format.

The original pathway diagram is drawn using the Adobe Illustrator. A pdf version of the pathway is available for download to our users.

9. Click on Supporting Evidence to download the evidence spreadsheet for detailed literature support evidence for each step of the pathway.

SUPPORT PROTOCOL 3: ORIENTATION TO THE VIP GENE PAGE

Very Important Pharmacogenes (VIPs) are structured summaries of key information for genes that are important for pharmacokinetic or pharmacodynamic effects of drugs. These in-depth annotated summaries include information on important variants of the gene, mapping information, haplotypes, population frequencies, phenotypes and interacting drugs. Supporting key PubMed references are included in each gene summary. These annotations are manually curated.

Necessary Resources

Hardware—Computer with an internet connection

Software—Any up-to-date browser will work

Files

No input files required

1. From the home page (<http://www.pharmgkb.org>), click the VIP gene icon.
2. This will lead to the page for all VIPs within PharmGKB, in alphabetical order. Click on View under the VIP Page column to go to the ABCB1 VIP gene summary. VIP pages can also be accessed from a gene page by clicking on the VIP tab.
3. On the ABCB1 VIP main page, one can find the gene name/symbol, summary, key PubMed IDs, associated pathways and drugs, and important haplotype information (Fig. 14.5.6).

The VIP gene page itself is constructed from a standard template. The list of contributors to the pathway is provided at the top, followed by links to any important variants, haplotypes, or splice variants that are associated with that gene. Below this information is the VIP summary. Included with every VIP summary are the HGNC gene name, common names or synonyms that frequently appear in the literature for this gene, an introductory paragraph, and key PubMed ID numbers that are associated with the information in the introductory paragraph. If applicable, the VIP page also contains links to PharmGKB pages for the drugs that this gene interacts with, the PharmGKB pathways that the gene appears in, and any phenotypes or diseases for which information is available. At the bottom of the main VIP page, there are links to the important variants, haplotypes, or splice variants that are associated with this gene. In order for a gene to qualify as a VIP gene candidate, it must have at least one variant of pharmacogenomic significance.

4. Click on Important Variants on the top right corner of the VIP gene page to go to the page with detailed information on important variants, external references, and their impact on drug responses (Fig. 14.5.7)

The VIP variant page is structured similarly to the main VIP page. The top of the VIP variant page contains the list of authors and links back to the main VIP summary, as well as any important haplotypes associated with this gene. Following below is a list showing how many important variants there are for each gene (e.g., there are three important variants for ABCB1). Each important variant has its own entry. This entry contains the HGNC name for the gene, a variant summary that is specific to that particular variant (in contrast to the general summary on the VIP gene page), key PubMed IDs that are associated with the variant summary, complete mapping information for the variants, links to relevant PharmGKB pages for the drugs, and links to phenotype datasets for the gene. The mapping information includes genomic position and accession number, a dbSNP unique identifier (number starts with rs), and its Golden Path position. If applicable, an mRNA and protein position and accession

number are also provided. Variant pages may also contain allele frequency tables that list a brief description of the population that has been studied, the number of subjects in that population, the allele frequency of the variant in question, and a link with the PMID number that opens a new browser window with that PubMed abstract. If the variant is part of a haplotype, then there is also a link included at the bottom of the page linking the user to the haplotype that contains the variant.

5. Click on Important Haplotype to go the page with detailed information on known haplotypes for gene of interest, related SNPs, and associated phenotype data files and their impact on drug responses

Haplotype pages are similar to the variant pages, and contain much of the same information. The differences are that there is no mapping information for haplotypes, and these pages also describe how many SNPs contribute to the formation of these haplotypes. A haplotype may be defined by only one SNP, as is the case with many of the CYP haplotypes. In this case, information on the haplotype page may be duplicated on the variant page and vice versa. A separate variant page for any CYP haplotype is included so that the mapping information for that position can also be incorporated. A haplotype page also contains a definitive publication or link to an external Web site, which will take the user to the source that was used to name the haplotype.

SUPPORT PROTOCOL 4: ORIENTATION TO PHARMGKB WEB SERVICES

PharmGKB web services enable our users to download a selected subset of data from PharmGKB via a Simple Object Access Protocol (SOAP) interface. Application programming interface (API) documentation and sample codes for any user who wishes to access portions of the database is available at <http://www.pharmgkb.org/home/projects/webservices/index.jsp>.

Necessary Resources

Hardware—Computer with an internet connection

Software—A user may create a Web services client program in any language of choice. PharmGKB provides Perl and Python clients to access Web services; detailed documentation is available at <http://www.pharmgkb.org/home/projects/webservices/index.jsp>.

Files—Sample codes (Perl and Python) is available for download to access genes, drugs, diseases and variants using PharmGKB accession numbers. Documentation is found at <http://www.pharmgkb.org/home/projects/webservices/README-perl.txt> or, <http://www.pharmgkb.org/home/projects/webservices/README-python.txt>

1. To use the Perl scripts, download and install the SOAP::lite module from <http://soaplite.com/> or <http://sourceforge.net/projects/soaplite> (available from CPAN)

API documentation: <http://cpan.uwinnipeg.ca/htdocs/SOAP-Lite/SOAP/Lite.html>

```
% cpan
cpan> install MIME::Parser
cpan> install SOAP::Lite
```

2. Alternatively, if the user is more familiar with writing or running Python scripts, download and install SOAPpy from:

http://sourceforge.net/project/showfiles.php?group_id=26590

in addition to the python module “fpconst” available from:

<http://cheeseshop.python.org/packages/source/f/fpconst/fpconst-0.7.2.tar.gz>

```
% cd fpconst-0.7.2
% python setup.py install
Then, install the SOAPpy module itself
% cd SOAPpy-0.12.0
% python setup.py install
```

3. For example, if Perl clients are used, “specialSearch.pl” will allow user to access the following types of data from PharmGKB (refer to Table 14.5.2 for special codes for specific data types):

For instance, typing “

```
perl specialSearch.pl 0
```

“ will output all genes with pharmacokinetic relevance at PharmGKB.

COMMENTARY

Background Information

PharmGKB began as the central data repository for the Pharmacogenetics Research Network (PGRN) and scientific community at large in 2000 (Giacomini et al., 2007; Long, 2007). It is designed to be a publicly available knowledge base with scientifically documented information connecting phenotypes to genotypes. Over the past eight years of development (funded by NIH), PharmGKB has grown to be an integrated resource that provides data on variants in genes, their relationship to drug response phenotype, the phenotype data, and curated knowledge in the forms of drug centered pathways, pharmacogene summaries (VIPs) and literature annotations (Hodge et al., 2007). PharmGKB currently houses variant data associated with more than 600 genes, greater than 2000 manually curated literature annotation, 52 drug-centered pathways, and 27 VIP gene summaries. Our comprehensive content makes it easier and faster for our users to access key pharmacogenomic information without repeating searches in multiple databases.

PharmGKB primary data comprises both genotype data and phenotype data. Initially, the data depository was seeded by data from the PGRN and mainly focused on a handful of

genes at a time. With the rapid advancement and widespread use of high throughput technology to measure gene variation and gene expression, the field of pharmacogenomics has evolved to explore a much larger set of genes, up to the whole genome. This also includes how variations in these larger gene sets work in concert to affect drug response. PharmGKB has expanded its capacity to accommodate large-scale high-throughput data which may involve large number of samples assayed across the entire genome. SNP array data can now be viewed and downloaded from the PharmGKB. PharmGKB also houses large data submissions from beyond PGRN. In 2006, Applied Biosystems posted genotype data for more than 220 drug response genes from four human populations (Caucasian, African American, Chinese, and Japanese) on PharmGKB. Allele frequencies for each of these populations were calculated and are available from the variant frequency report. PharmGKB is also the central repository for the International Warfarin Pharmacogenetics Consortium (IWPC). (<http://www.pharmgkb.org/views/project.jsp?pid=56>). The goal of this consortium is to create a merged international dataset (including more than 5,700 patients) in order to develop the best strategy for predicting the therapeutic dose of warfarin.

In addition to our data mission, PharmGKB is curating pharmacogenomic knowledge, including summarizing drug-centered pathways, annotating very important pharmacogenes (VIPs) and primary literature. Unlike other pathway resources—e.g., KEGG (*UNIT 1.12*), Reactome (*UNIT 8.7*), Biocarta, GenMAPP (*UNIT 7.5*) that primarily focus on physiological processes, PharmGKB is the only resource that focuses on drug-centered pathways, particularly pharmacokinetic (PK) pathways. This effort is valuable to the scientific community as our pathways enable researchers to conduct in-depth analysis of various forms of experimental data within the framework of curated drug response pathways. Our pharmacokinetics (PK) pathways describe candidate genes involved in the absorption, distribution, metabolism and excretion of a given drug, while the pharmacodynamic (PD) pathways illustrate the physiological effects of the drug, its mechanism of action and possible side effects. Currently, there are 39 interactive drug-centered pathways created in collaboration with experts in the pharmacogenomic area. PharmGKB pathways have been widely quoted by the scientific community for their unique content (Mangravite et al., 2006; Scripture and Figg, 2006). The VIP gene summaries are another unique knowledge-rich feature provided by PharmGKB for key genes that are involved in modulating drug response. Each VIP summary is constructed using a structured template and includes detailed information about a given gene, including its important polymorphisms, haplotypes, phenotypes, and complete mapping information. An allele frequency table may also be included if the specific variant is studied extensively in different populations. VIP summaries are encyclopedia-like encapsulations for genes that require tremendous amounts of manual curation. They can potentially save scientists countless hours of time in their own literature mining process, which can be tedious, repetitive and time consuming. To keep our pathways and VIPs current, PharmGKB updates them every two years to incorporate any new interactions or correct any erroneous information that is being displayed.

PharmGKB provides a wealth of information to facilitate the design of a pharmacogenetics study such as identifying genetic markers for a patient's response to a therapeutic agent. A

scientist designing the study can use PharmGKB in the following manner to pick the best candidates genes and variants from our integrated knowledge base.

Identify candidate genes important for pharmacokinetics or -dynamics of the drug used in the study—If the pathway for the specific drug is available through PharmGKB, this is the best place to start looking for the candidate genes. The genes on our drug-centered pathway are known to be involved in the disposition or mechanism of action of the drug, and the user can click on each gene in the pathway to delve down to the detailed variant information associated with that gene. If no pathway is currently available for the drug of interest, the scientist can first perform a search for the drug in PharmGKB, open the drug page and then go to the section on “*Related Genes From Literature*” to find candidate genes that are implicated in drug response as well as their literature evidence to decide on which genes to choose for the study.

Find functional variants for the candidate genes chosen—If there is a PharmGKB VIP page available for the gene, the VIP will identify the important variants and haplotypes for the gene of interest. Alternatively, the scientist can go to PharmGKB variant page and browse through the variant table which lists all the variants for the gene, their genomic position, functional role, frequency and assay type. SNPs that reside in the exon or promoter regions of the gene, and SNPs that lead to changed amino acid composition, inactive protein or changed expression of the gene are good candidates to be included for the study. Annotations for variants that have been studied for phenotypic consequences are tagged with the star system as discussed in support protocol 1.

Determine if the population frequencies for the chosen variants are desirable—This step will further screen out SNPs that are may be to rare in the population that will be included in the study. The frequency information can be found in the frequency column on PharmGKB variant table. Clicking on the frequency value displays the breakdown of frequencies by racial categories.

Find assay and primer information for the chosen variants—Clicking on the nucleotide changes in *Variant column* of the PharmGKB variant table will allow the user to find information such as assay methods and primers. For instance, if the Taqman assay was used to genotype a specific drug metabolizing enzyme variant, PharmGKB provides a direct link to ordering information at Applied Biosystems to help the user identify the material required for the study. By iterating through these steps, a scientist can compile a short list of candidate genes and SNPs that can be used in a study to identify genetic markers that might explain and predict the efficacy and adverse effect profiles of the drug of interest.

Pharmacogenomics is a rapidly evolving field with many unmet challenges in translating the scientific findings in pharmacogenomics to clinical practice. However, the increasing understanding of how a person’s genetic makeup can influence his or her response to drugs provides the opportunity to improve the drug development process and provide more effective and safer therapy for individual patients. PharmGKB will continue its efforts to aggregate, integrate and annotate the latest findings in pharmacogenomic research, and provide tools and context to catalyze scientific discoveries.

Critical Parameters and Troubleshooting

PharmGKB is designed to be a valuable resource for both expert researchers in the pharmacogenomics field, as well as for novice users and the general public. PharmGKB's homepage prominently displays the information that our users are looking for most frequently with a distinct icon system to represent different data types and knowledge. Typical searches conducted at PharmGKB are for information about drugs, genes, diseases and pathways. Searches can be conducted using the Web-search engine-like search box. If too many search results are returned, users can narrow the search with more specific terms. Alternatively, a user can use our Search tab and simple query to limit the domain of search to specific area of interests (e.g., genes with genotype data; the relevant literature on drug X; diseases with PharmGKB primary data). If the user encounters difficulties in finding information of interest, using alternative names, partial names, or a loosening of the search criteria is suggested. Alternatively, if nothing is returned under the "database search" tab, the user should look for results under the "website search" tab as PharmGKB has full-text indexing to allow our users to search across the entire website.

We welcome all feedback regarding the PharmGKB. Questions and concerns can be sent to feedback@pharmgkb.org. Our scientific staff will respond to your inquiry within 48 hr.

Acknowledgments

PharmGKB is supported by the NIH/NIGMS Pharmacogenetics Research Network (PGRN) (U01GM61374). The authors thank the entire PharmGKB team (<https://www.pharmgkb.org/home/team.jsp>) that has contributed to the development of PharmGKB.

Literature Cited

- Eyre TA, Ducluzeau F, Sneddon TP, Povey S, Bruford EA, Lush MJ. The HUGO Gene Nomenclature Database, 2006 updates. *Nucleic Acids Res.* 2006; 34:D319–D321. [PubMed: 16381876]
- Giacomini KM, Brett CM, Altman RB, Benowitz NL, Dolan ME, Flockhart DA, Johnson JA, Hayes DF, Klein T, Krauss RM, Kroetz DL, McLeod HL, Nguyen AT, Ratain MJ, Relling MV, Reus V, Roden DM, Schaefer CA, Shuldiner AR, Skaar T, Tantisira K, Tyndale RF, Wang L, Weinshilboum RM, Weiss ST, Zineh I. The pharmacogenetics research network: From SNP discovery to clinical drug response. *Clin Pharmacol Ther.* 2007; 81:328–345. [PubMed: 17339863]
- Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005; 33:D514–D517. [PubMed: 15608251]
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004; 32:D258–D261. [PubMed: 14681407]
- Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y. JSNP: A database of common gene variations in the Japanese population. *Nucleic Acids Res.* 2002; 30:158–162. [PubMed: 11752280]

- Hodge AE, Altman RB, Klein TE. The PharmGKB: Integration, aggregation, and annotation of pharmacogenomic data and knowledge. *Clin Pharmacol Ther.* 2007; 81:21–24. [PubMed: 17185992]
- Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L. Reactome: A knowledgebase of biological pathways. *Nucleic Acids Res.* 2005; 33:D428–D432. [PubMed: 15608231]
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 2004; 32:D277–D280. [PubMed: 14681412]
- Klein TE, Altman RB. PharmGKB: The pharmacogenetics and pharmacogenomics knowledge base. *Pharmacogenomics J.* 2004; 4:1. [PubMed: 14735107]
- Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW, Stanke M, Smith KE, Siepel A, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pedersen JS, Hsu F, Hinrichs AS, Harte RA, Diekhans M, Clawson H, Bejerano G, Barber GP, Baertsch R, Haussler D, Kent WJ. The UCSC genome browser database: Update 2007. *Nucleic Acids Res.* 2007; 35:D668–673. [PubMed: 17142222]
- Letovsky SI, Cottingham RW, Porter CJ, Li PW. GDB: The Human Genome Database. *Nucleic Acids Res.* 1998; 26:94–99. [PubMed: 9399808]
- Long RM. Planning for a national effort to enable and accelerate discoveries in pharmacogenetics: The NIH Pharmacogenetics Research Network. *Clin Pharmacol Ther.* 2007; 81:450–454. [PubMed: 17339874]
- Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Res.* 2007; 35:D26–3D1. [PubMed: 17148475]
- Mangravite LM, Thorn CF, Krauss RM. Clinical implications of pharmacogenomics of statin treatment. *Pharmacogenomics J.* 2006; 6:360–374. [PubMed: 16550210]
- Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 2007; 35:D61–D65. [PubMed: 17130148]
- Safran M, Solomon I, Shmueli O, Lapidot M, Shen-Orr S, Adato A, Ben-Dor U, Esterman N, Rosen N, Peter I, Olender T, Chalifa-Caspi V, Lancet D. GeneCards 2002: Towards a complete, object-oriented, human gene compendium. *Bioinformatics.* 2002; 18:1542–1543. [PubMed: 12424129]
- Scripture CD, Figg WD. Drug interactions in cancer therapy. *Nature Rev.* 2006; 6:546–558.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* 2001; 29:308–311. [PubMed: 11125122]
- Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 2006; 34:D668–D672. [PubMed: 16381955]
- Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. The Universal Protein Resource (UniProt): An expanding universe of protein information. *Nucleic Acids Res.* 2006; 34:D187–D191. [PubMed: 16381842]

PharmGKB
The Pharmacogenetics and Pharmacogenomics Knowledge Base

Search PharmGKB: ?

Home Search Submit Help PGN Contributors Sign In Feedback

PharmGKB curates information that establishes knowledge about the relationships among drugs, diseases and genes, including their variations and gene products. Our mission is to catalyze pharmacogenomics research.

Click for more information

genotyped genes	variants of interest	drugs	pathways	diseases	important PGx genes	SNP arrays	downloads & services
608	300	546	54	543	36	2	

Search PharmGKB: ?
e.g., a gene ("NOS3"), drug ("ace inhibitors") or disease ("ALL")

PGx Information Flow

Items in the flow chart below are clickable:

What's New?

- New PharmGKB Initiative: Variants of Interest (Click on Second Home Page Icon to View!)
- For Submitters: Status of Submission to dbSNP is shown on your group's submission page!

See the [archives](#) for more.

Curator's Favorite Papers

- Genetic variation and gene expression relationship for pharmacogenes [FA](#) [GN](#)
- Folate-remedial MTHFR variants in humans [FA](#) [GN](#)

Updated 6/16/08. See the [archives](#) for more.

Sign In

User Id

Password

Figure 14.5.1.
The PharmGKB home page (<http://www.pharmgkb.org>)

PharmGKB
The Pharmacogenetics and Pharmacogenomics Knowledge Base

Search PharmGKB:

Home Search Submit Help PGRN Contributors Sign In Feedback

VKORC1
vitamin K epoxide reductase complex, subunit 1

Overview VIP Variants Datasets Pathways Curated Publications Downloads/Cross-references

Overview

PharmGKB Accession Id: PA133787052

Alternate Names: VKCFD2; Vitamin K-dependent clotting factors, combined deficiency of, 2; phyloquinone epoxide reductase; vitamin K dependent clotting factors deficiency 2; vitamin K1 epoxide reductase (warfarin-sensitive)

Alternate Symbols: EDTP308; FLJ00289; IMAGE3455200; MGC2694; MST134; MST576; UNQ308; VKCFD2; VKOR

Details

Cytogenetic Location: chr16 : [p11.2](#)

GP mRNA Boundary[†]: chr16 : 31003777 - 31012676

GP Gene Boundary[†]: chr16 : 31003777 - 31018002

Strand: minus

OMIM Phenotype: Vitamin K-dependent clotting factors, combined deficiency of, 2, Warfarin resistance

Product Name: Vitamin K-dependent clotting factors, combined deficiency of, 2, phyloquinone epoxide reductase, vitamin K dependent clotting factors deficiency 2, vitamin K epoxide reductase complex, subunit 1, vitamin K epoxide reductase complex, subunit 1 isoform 1, vitamin K epoxide reductase complex, subunit 1 isoform 2, vitamin K1 epoxide reductase (warfarin-sensitive)

[†] The mRNA boundaries are calculated using the gene's default feature set from NCBI, mapped onto Golden Path. The gene boundaries are set by PharmGKB by expanding the mRNA boundaries to allow for upstream and downstream regulatory regions.

Figure 14.5.2.
Example of PharmGKB Gene page (VKORC1, vitamin K epoxide reductase complex, subunit 1)

Figure 14.5.3.
Example of PharmGKB phenotype data.

Figure 14.5.4.
Example of PharmGKB Variant Gene Page (VKORC1) with variant browser on the top and variant table below

Figure 14.5.5.
Example of PharmGKB pathway (Irinotecan pathway)

Annotated PGx Gene Information for ABCB1

Submitted by: Jason Gow, Leslie Chinn, PMT Group		Jump To: ☐ Important Variants ☐ Important Haplotype ☐ All Annotated Genes
Submitted date: August 11, 2008 Feedback		

Gene HGNC Name:	ABCB1
Gene Common Name:	MDR1
Introductory Information:	<p>The ABCB1 gene is a member of the ATP-binding cassette (ABC) transporter superfamily. The gene is also known as the multidrug-resistance 1 (MDR1) gene and its protein product is called P-glycoprotein (P-gp). ABCB1 is expressed in barrier and excretory tissues that have protective or elimination roles, such as the intestines, liver, kidney, blood-brain barrier, and placenta. ABCB1 over-expression in tumors has been implicated in multidrug resistance to cancer chemotherapeutic agents. Located on the plasma membrane, P-gp can transport a wide range of endogenous and exogenous compounds out of the cell [PMID: 10331089]. Substrates of P-gp include anticancer agents, cardiac drugs and HIV protease inhibitors. Studies have demonstrated the existence of interindividual variability in P-gp expression and function as well as clinical phenotypes related to P-gp [PMID: 7473127].</p> <p>ABCB1 is located on chromosome 7q21.1 and was originally cloned by Riordan <i>et al.</i> in 1985 [PMID: 2863759]. The gene consists of 29 exons spanning nearly 200 kb of genomic DNA; however only 27 exons code for P-gp. There are multiple transcriptional start sites but the primary start site produces an mRNA transcript of 4360 bp that contains 28 exons. Currently, there is no evidence for the existence of splice variants. P-glycoprotein has 1280 amino acids, 12 transmembrane domains and two ATP-binding domains [PMID: 10331089].</p> <p>ABCB1 has approximately 116 polymorphic sites in Caucasians and 127 in African Americans with a minor allele frequency greater than 5% (www.hapmap.org [PMID: 12893986, 11503014, 10716719]). Some of the most commonly studied variants are 1236C>T, 2677G>A/T and 3435C>T and the most commonly studied haplotype involves the 1236, 2677 and 3435 SNPs. There are many other ABCB1 variants such as -129C>T (5'-UTR), 61A>G (Asn21Asp) and 1199G>A (Ser400Asn) that have been studied <i>in vivo</i> and <i>in vitro</i>. To date, there is no clear consensus on the impact of any of these variants on drug disposition, response or toxicity. For the most commonly studied phenotypes, i.e. intestinal absorption and CNS penetration, discordant results have been published and appropriately powered studies are still needed to clarify these results. Helpful review papers have been written that summarize the substantial body of work covering ABCB1 polymorphisms and their potential impact on cellular and clinical phenotypes [PMID: 12505329, 12406646, 14749689, 15212152].</p>
Key Pubmed IDs:	PMID: 12505329 , 12406646 , 14749689 , 15212152 , 10331089 , 2863759 , 7473127 , 11503014 , 12893986 , 10716719
Key Pathways:	Etoposide , Irinotecan

Figure 14.5.6.
Example of VIP gene page (ABCB1)

Important Variant Information for ABCB1

Submitted by: Jason Gow, Leslie Chinn, [PMT Group](#)

Submitted date: August 11, 2008

[Feedback](#)

Jump To:

[ABCB1 Overview](#)

[Important Haplotypes](#)

[All Annotated Genes](#)

There are **Three Important Variants** for ABCB1.

1. [1236C>T](#) (rs1128503)

2. [2677G>A/T](#) (rs2032582)

3. [3435C>T](#) (rs1045642)

1. 1236C>T (rs1128503)

Gene HbC Name:	ABCB1
Related Haplotypes:	ABCB1*13
Genomic Variant & GenBank ID:	84224C>T on AC005068
MRI& Variant & GenBank ID:	1654C>T on NM_000927.3
Protein variant & GenBank ID:	NA
dbSNP RSC:	rs1128503
dbSNP SSR:	ss35072107
GoldenPath Position:	Chr7: 86,824,252
Phenotypes/Diseases:	Marginal association with CD4 cell recovery after protease inhibitor treatment [PMID: 15651752]
Publicly available DNA source (e.g. Coriell line) containing homozygous reference allele:	Coriell NA17206 (P#A126721672)
Publicly available DNA source (e.g. Coriell line) containing heterozygous reference allele:	Coriell NA17204 (P#A126721670)
Publicly available DNA source (e.g. Coriell line) containing homozygous minor allele:	Coriell NA17208 (P#A126721674)
Summary text	This SNP is positioned at 1236 with respect to the translation start site. The body of work studying the effect of the 1236 C>T synonymous SNP is fairly small, with no conclusive findings to date. The majority of reports involving the 1236 SNP consider this polymorphism in its naturally occurring haplotype *13 (see below). This SNP is in exon 12.
Key PubMed IDs:	PMID: 15651752 , 12960109

Figure 14.5.7.
Example of important variant page for VIP genes (ABCB1)

Curr Protoc Bioinformatics. Author manuscript; available in PMC 2014 September 03.

Table 14.5.1

Description of Menu Tabs

Tab	Description
Home	The front page where we highlight our knowledge and data content, mission, contact info and registration
Search	The main search page where we can either search by free text, user- canned queries, or browse information by domain
Submit	The section that describes how a user can submit genotype, phenotype, pathway or literature data
Help	An extensive list of background information, downloads, educational as well as technical references
PGRN	Lists all members involved in the NIH Pharmacogenetics Research Network, their research interests and submissions to PharmGKB
Contributors	The section where people are listed who have contributed data to PharmGKB
My PharmGKB	The section for our registered users to view their profile, submission and website statistics

Table 14.5.2

Codes for Use With specialSearch.pl

Code	Data Type
0	Genes with pharmacokinetic (PK) significance
1	Genes with pharmacodynamic (PD) significance
2	Genes with PharmGKB variant data
3	Genes with PK variants
4	Genes with PD variants
5	Drugs with supporting information
6	Diseases with supporting information
7	Phenotype datasets
8	Pathways with PGx significance
9	Annotated publications describing relationships between genes, drugs and diseases
10	Literature annotations, pathways and phenotype datasets annotated with a pharmacokinetics (PK) COE
11	Literature annotations, pathways and phenotype datasets annotated with a pharmacodynamics (PD) COE
12	Literature annotations and phenotype datasets annotated with a clinical outcome (CO) COE