



HAL
open science

Databases in the area of Pharmacogenetics

Sarah C Sim, Russ B Altman, Magnus Ingelman-Sundberg

► **To cite this version:**

Sarah C Sim, Russ B Altman, Magnus Ingelman-Sundberg. Databases in the area of Pharmacogenetics. *Human Mutation*, 2011, 32 (5), pp.526. 10.1002/humu.21454 . hal-00629068

HAL Id: hal-00629068

<https://hal.science/hal-00629068>

Submitted on 5 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Databases in the area of Pharmacogenetics

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0559.R1
Wiley - Manuscript type:	Review
Date Submitted by the Author:	20-Dec-2010
Complete List of Authors:	Sim, Sarah; Karolinska Institutet, Department of Physiology and Pharmacology Altman, Russ; Stanford University, Department of Bioengineering Ingelman-Sundberg, Magnus; Karolinska Institutet, Department of Physiology and Pharmacology
Key Words:	Pharmacogenetics, database, personalized medicine, adverse drug reaction, cytochrome P450, transporter, polymorphism

SCHOLARONE™
Manuscripts

Databases in the area of Pharmacogenetics

Sarah C Sim^{1/*}, Russ B Altman^{2/} and Magnus Ingelman-Sundberg^{1/}

^{1/} Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, SE-171 77 Stockholm, Sweden; ^{2/} Department of Bioengineering, Stanford University, Stanford, California 94305, USA

*Correspondence to Sarah C Sim (Email: Sarah.Sim@ki.se, Phone +46 8 524 862 39, Fax +46 8 33 73 27)

ABSTRACT

In the area of pharmacogenetics and personalized health care it is obvious that databases providing important information of the occurrence and consequences of [variant](#) genes encoding drug [metabolizing](#) enzymes, drug transporters, drug targets and other [proteins](#) of importance for drug response or toxicity, are of critical value for scientists, physicians and [industry](#). The [primary](#) outcome of the pharmacogenomic field is the identification of biomarkers that can predict drug toxicity and drug response, thereby individualizing and improving drug treatment of patients. [The drug in question and the polymorphic gene exerting the impact are the main issues to be searched for in the databases. Here, we review the databases that provide useful information in this respect, of benefit for the development of the pharmacogenomic field.](#)

Key words: [Pharmacogenetics, database, personalized medicine](#), adverse drug reaction, cytochrome P450, transporter, polymorphism

Deleted: variants of

Deleted: metabolising

Deleted: genes

Deleted: for drug development.

Deleted: drug in question and the polymorphic gene exerting the impact are the main issues to be searched for in the databases. Here, we review the databases in the field that provide useful information in this respect. One important

Deleted: Website, webpage, home page

1. Introduction

There is a rapid development in genetic analyses and functional genomics covering next generation sequencing methods and versatile arrays that rapidly [analyze](#) the most important mutations and gene copy number variations in the genome. Based on this progress, it becomes increasingly evident that much [but not all](#) of the drug therapy can be individualized based on genome sequence variation. These efforts can result in pharmacogenomic drug labels that advise the physician to use a genetic test prior to drug prescription. This test can be mandatory, recommended or for information purposes only (see [\(Sim and Ingelman-Sundberg, 2011\)](#) for a review). The genetic variation of interest mainly concerns genes encoding drug metabolizing enzymes, drug transporters, drug targets and human leukocyte antigens (HLA). [Only](#) 30-60 % of patients respond to common drug therapy using e.g. antidepressants, statins, and antipsychotics ([Spear et al., 2001](#)). Furthermore, adverse drug reactions (ADRs) are [an](#) increasing [problem](#) (Moore et al., 2007) and [apparently](#) cost society as much as the drug therapy itself ([cf. Eichelbaum et al., 2006](#)). In addition, meta-analyses by Lazarou *et al.* (1998) revealed consequences of [more than](#) 100,000 deaths annually [caused by](#) ADRs in the US, [and](#) [Davies et al. \(2009\) have documented an increased length of hospitalization for more than 25% of patients experiencing ADRs in the UK. Also, ADRs are shown to](#) occur in 16 % of hospitalized patients, frequently with severe consequences (Davies et al., 2009), [and](#) ADRs have in several studies been shown to be responsible for about 7 % of hospitalizations [in the average](#) [population](#) (Lazarou et al., 1998; Davies et al., 2009), [20 % of readmissions to the hospital](#)

Deleted: analyse

Deleted: the

Deleted: In support of this development is the fact that only

Deleted: cf. Eichelbaum

Formatted: Font: Italic

Deleted: 2006

Deleted: in frequency

Deleted: the

Deleted: .

Deleted:

Deleted: and prolonged hospital stay due to

Deleted: . In Europe, a prospective study showed that ADRs

Deleted:). Furthermore,

Deleted:) in the average population

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

([Davies et al., 2010](#)) and more than 30 % among elderly (Paul et al., 2008). These data emphasize the need of updated databases that can provide fruitful guidance to both scientists, physicians, regulatory agencies and industries [to cope with this major problem in human health](#).

Deleted: .

2. The Pharmacogenomics Knowledge Base (PharmGKB)

The Pharmacogenomics Knowledgebase (PharmGKB, <http://www.pharmgkb.org/>) was established in 2000, and strives to provide information on how human genetic variation impacts drug response. Stable government-funded efforts such as Genbank, Gene Expression Omnibus, and dbGAP from the National Center for Biomedical Information (NCBI) at the National Institutes of Health (NIH) in the US are the preferred databases for [archiving](#) data, whereas curation of genes (or groups of related homologous genes) and their variations are best handled by locus-specific databases (LSDBs) such as the Human Cytochrome P450 (CYP) Allele Nomenclature Website ([see section 3](#)), and others focusing on transporters, receptors, and kinases. [In](#) contrast, PharmGKB focuses on four main activities: (1) curation of published literature [in respect](#) to relationships between genes, their variations, and drug response phenotypes, (2) creation of "Very Important Pharmacogene" (VIP gene) summaries that highlight specific pharmacogenomic significance of genes that are very important for drug response, (3) creation of pathway diagrams focusing on drug response (pharmacodynamics, PD) and drug metabolism (pharmacokinetics, PK) pathways, and finally (4) [assistance](#) to organize and lead data-sharing consortia within pharmacogenomics where multiple centers combine data and expertise to answer questions that cannot be answered by any center alone (usually because of statistical power and smaller data sets).

Deleted: ,

Deleted: achieving

Deleted: ,

Deleted: By

Deleted: as a knowledgebase

Deleted: the

Deleted: the

Deleted: the

Deleted: based on the information gathered in (1) and (2),

Deleted: PharmGKB has helped

Deleted: Here, we review each of these areas of activity and provide an update.

2.1 Curation of published literature

The core activity of PharmGKB curators is to survey the literature and capture reports of human genetic variation that impacts drug response phenotypes. These phenotypes may occur at the

molecular (e.g. binding), cellular (e.g. expression), tissue (e.g. drug transformation) or

organismal level (e.g. disease, symptoms), and all annotations are encoded with respect to

these categories to assist in searches. The PharmGKB has a list of journals publishing

pharmacogenomics observations, and also broadly searches the literature on a weekly basis to

find associations between genes, their variations, drugs, and phenotypes. The PharmGKB uses

the Human Genome Nomenclature Committee recommendations for genes and variations. Also,

they use a combination of the [RcNorm](#) drug list with some modifications for drug, and [Medical](#)

[Subject Headings \(MeSH\)](#) and [Unified Medical Language System \(UMLS\)](#) to encode phenotypes.

The result of this curation activity is currently about 2500 annotations that associate particular

human variations (mostly SNPs, but also including insertions or deletions (indels), copy number

variations, structural changes, and haplotypes) with drug-related phenotypes. These

annotations were recently used to create a "pharmacogenomic annotation" of a complete

human genome sequence (Ashley et al., 2010). Interestingly, there were nearly 100 drugs for

which risks or benefits could be predicted for this individual. Many of them were "low

confidence" based on the level of published evidence, but several of them would have some

clinical utility. The application of the PharmGKB annotations to this genome was manual, and so

current efforts are aimed at representing the basic variant annotations and their potential

clinical consequences in a more structured manner that will allow more automated application

to individual genomes.

Deleted: ¶

Deleted: any

Deleted: on

Deleted: basistobe able

Deleted: RXNORM

Deleted: MESH

Deleted: ; Evans 1991

Deleted: at the same time

2.2 Writing VIP gene summaries

As the PharmGKB curators accumulate individual variant annotations, the genes may appear as either (1) repeatedly involved in drug response because of their role in general drug pharmacokinetics (e.g. cytochrome P450s, transporters, transferases) or (2) having an important role in the action of a single drug or class of drugs (e.g. pharmacodynamic drug targets). When the amount of information for these genes reaches a level that the curators find to be sufficiently large, they can designate these genes as VIP genes¹. The PharmGKB curators prioritize these genes for short written summary reviews of their role in drug response, and the key variations in these genes that have been shown to modulate drug response. These reviews are re-examined every two years or more frequently if the pace of new knowledge generation is rapid, and are written in a templated manner that includes names and alternate names, introductory information, key publications, key pathways, known drugs/substrates, associated phenotypes/diseases and summaries of key variants with their associated effects. There are currently 43 VIP genes on the PharmGKB website, and their summaries are published regularly in a cooperative venture with the *Pharmacogenetics & Genomics* journal in order to draw attention to them and to provide a citation for their usage (Eichelbaum et al., 2009; Mi et al., 2010).

Deleted: ,

2.3 Creating pathways of drug response and metabolism

¹Very Important Pharmacogene (VIP gene) denotes genes that have a documented role in modulating the response to one or more drugs

Deleted: Pharmacogenes

Deleted: genes) denote

1
2
3 In addition to the VIP gene summaries, curators also create annotated graphical pathways
4 depicting the relationships of genes and drugs. These pathways are the most popular feature at
5 PharmGKB. Each pathway is the result of a PharmGKB curator gathering all information
6 available for the drug of interest (for example, by collecting the basic annotations and gene-
7 drug relationships discussed above), and assembling a pathway summary. These pathways are
8 often created collaboratively with [experts](#) working on these systems and [they include a down-
9 loadable](#) graphical depiction that is linked to the individual genes and small molecules depicted.
10 It also includes [supporting data and references](#) upon which the pathway is based. As with the
11 VIP genes, the pathways are also published regularly in the *Pharmacogenetics & Genomics*
12 journal (Eichelbaum et al., 2009; Sangkuhl et al., 2010). In addition to PharmGKB-generated
13 pathways, PharmGKB links to pathways from external resources, and they are made available to
14 augment the PharmGKB pathways.

Deleted: , drawing the most internet "hits."

Deleted: high quality

Deleted: expert collaborators

Deleted: the yinclude

Deleted: downloadable

Deleted: associated

Deleted: and provide "one stop shopping" for users interested in drug-related pathways.

2.4 Data sharing consortia

30 A somewhat unexpected activity for PharmGKB has been the convening of data sharing
31 consortia. The mission of PharmGKB is to promote the creation and sharing of data about
32 pharmacogenomics, and [as a trusted third party](#), PharmGKB is in a good position to bring
33 together groups of investigators interested in common scientific questions. After signing a
34 memorandum of understanding that outlines the responsibilities and privileges of joining the
35 consortium (and protects the rights of data generators to publish, as well as the rights of the
36 consortium members to see data and participate in its analysis), the PharmGKB accepts data
37 sets, brings them into a common format, and then participates in the statistical analysis to

Deleted: the curators have outstanding skills at data organization, representation, aggregation and integration. As

Deleted: that are difficult to answer individually because of small sample sizes.

Deleted: of these data

1
2 answer one or more shared scientific questions. The first example of a PharmGKB [consortium](#)
3
4 was the International Warfarin Pharmacogenetics Consortium (IWPC) which brought together
5
6 21 groups internationally to create a globally-relevant warfarin dosing algorithm, as well as an
7
8 analysis of inter-ethnicity differences in warfarin dosing (Klein et al., 2009; Limdi et al., 2010).
9
10 Other consortia have subsequently been established addressing the discovery of new genetic
11
12 variations relevant to warfarin, the role of CYP2D6 variations in tamoxifen response, and the
13
14 role of CYP2D6 in selective serotonin reuptake inhibitor (SSRI) response. A list of consortia and
15
16 their members are available on the "Contributors" tab of PharmGKB.
17
18
19
20

Deleted: consortia

21 2.5 The PharmGKB website and other activities

22
23 In addition to a global search capability, users of PharmGKB are able to search for particular
24
25 kinds of information in five more focused search boxes: genes, variants, pathways, drugs, and
26
27 diseases. These search boxes can accept any text, but will only return results of one of these 5
28
29 specific types. In addition, the home page provides the PharmGKB curators "Favorite Papers"
30
31 that are updated every two weeks, and [are](#) available as an archive over the last four years. The
32
33 home page also offers an automatically created pharmacogenomics news feed, access to
34
35 tutorial information, introductions to pharmacogenetics, and access to tools.
36
37
38
39

Deleted: the

Deleted: (for

Deleted:).

Deleted: useful

40 Current new efforts at PharmGKB [focus on](#):

Deleted: to add new capabilities to help pharmacogenetic investigators connect genetic associations to their underlying mechanisms. These include

- 41
42 • The development of automated natural language parsing tools to extract annotation
43
44 information from the literature automatically (Coulet et al., 2010; Garten and Altman 2010;
45
46 Garten and Altman 2009).
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- The addition of [chemoinformatics](#), such as drug structure similarity metrics to help predicting new gene-drug interactions (Garten et al., 2010; Hansen et al., 2009).
- The creation of systems biology methods to look at drug response with a more network-based view of the genes involved in producing complex drug responses (Hansen et al., 2009; Tatonetti et al., 2009).

Deleted: chemoinformatic information

Deleted: not simply on a single gene basis, but

3. The Human Cytochrome P450 (CYP) Allele Nomenclature Website

Cytochrome P450 enzymes are responsible for 75-80 % of all phase I-dependent metabolism, [resulting in more polar metabolites that can be further conjugated in Phase II metabolism](#), and for 65-70 % of the clearance of clinically used drugs (Bertz and Granneman 1997; Evans and Relling 1999). [CYP enzymes](#) are encoded by 57 *CYP* genes and divided into families, of which 1-3 (CYP1-3) are generally involved in the metabolism of exogenous substances including drugs, whereas [members of](#) other families mainly are involved in metabolism of endogenous compounds. Variation in *CYP* genes results in four different phenotypes classically defined as ultrarapid, extensive, intermediate and poor [metabolizers](#).

Deleted: (CYP)

Deleted:), and they

Deleted: active

Deleted: . CYP enzymes are

Deleted: families

Characterization of the genetic polymorphism of CYP enzymes provides a basis for personalized medicine composing [drug dose adjustment and selection of drug type](#) (Ingelman-Sundberg 2004; Ingelman-Sundberg et al., 2007; Kirchheiner et al., 2004; Pirmohamed and Park 2001; Weinshilboum 2003).

Deleted: metabolisers.

Deleted: of

Deleted:), which includes improved response and avoidance of ADRs.

To unite and summarize efforts focusing on the characterization of CYP polymorphism, the Human Cytochrome P450 Allele [\(CYP-allele\)](#) Nomenclature Committee was established in 1999, and the efforts resulted in agreed *CYP* nomenclature guidelines and the launching of the *CYP*-

Deleted: (CYP-allele)

allele website (<http://www.cypalleles.ki.se/>). The main purpose of the *CYP*-allele website is the management of new allele designations based on recognized nomenclature guidelines, facilitation of rapid publication, as well as providing a [readily available](#) summary of alleles and their associated effects (Ingelman-Sundberg and Oscarson 2002; Oscarson and Ingelman-Sundberg 2002; Sim and Ingelman-Sundberg [2006](#); Sim and Ingelman-Sundberg [2010](#)).

At present, 29 polymorphic CYPs are included on the *CYP*-allele website, [with the addition](#) in 2008 [of](#) NADPH cytochrome P450 reductase (POR) (Sim et al., 2009) due to its elementary function [as electron donor](#) to CYP enzymes. Every gene has its own webpage, listing the associated alleles and their consequences on the genetic, molecular, *in vitro* and *in vivo* [levels](#), all together with relevant references and links to the NCBI dbSNP database.

The nomenclature system designates numbers (such as *CYP2B6*4*) for alleles that have been thoroughly sequenced and well characterized in terms of linkage of the [functionally relevant](#) SNP with other polymorphic sites. [Functionally relevant](#) in this aspect is regarded as a variation causing measurable consequences such as amino acid substitutions, translation terminations, [or](#) splice defects. At a minimum, the [open reading frame](#) must [have been](#) sequenced. All sequence variations are numbered [based on](#) A in the translation initiation codon as the reference nucleotide (+1). Additional letter-designated sub-alleles [of the same number](#) (such as *CYP2B6*4B*) [contain the same functional SNP but additional](#) sequence variations [not causing functional effects](#). A combination of [functionally relevant](#) variants are designated a new allele number due to the potential of combinatorial effects, unless one of the variant is deleterious and thus defines the allele name. [As an example, CYP2C19*2 carries a SNP causing a splice defect of the transcribed RNA that is detrimental for enzyme function. This splice defect](#)

Formatted: Font: Italic

Deleted: as

Deleted: 2010

Deleted: 2006

Deleted: and

Deleted: ,

Deleted: was included as well (cf.

Deleted:))

Deleted: to transfer electrons

Deleted: to facilitate their catalytic activity.

Deleted: level

Deleted: allele

Deleted: causative

Deleted: "Causative"

Deleted: functional

Deleted: ,

Deleted: , differential transcription rates etc.

Deleted: full coding region

Deleted: be

Deleted: , although flanking regions and introns would preferentially be covered as well, especially when a gene has shown to carry important variations in these regions.

Deleted: with

Deleted: contains non-causative

Deleted: in addition to the causative one that defines the allele. Such sub-alleles are named when they have been characterized together with the mother allele (e.g. *CYP2B6*4A*), but are

Deleted: designated thereafter due to the uncertainty whether they compose new alleles or just additional coverage by the sequencing method.

Deleted: causative

Deleted: that can also be found alone

Deleted: whereby the combination allele will be designated a sub-allele name.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(c681G>A) exists in different combinations with other SNPs causing amino acid substitutions and the combination alleles have thus received allele names CYP2C19*2A-D. The effect of CYP2C19 genotype on clopidogrel (Plavix®) antiplatelet treatment of cardiovascular disease has attained much focus since CYP2C19 is responsible for bioactivation of the drug. A recent meta-analysis of 9 studies and almost 10,000 patients showed a significantly increased risk of adverse cardiovascular events in carriers of defective (mainly CYP2C19*2) alleles due to impaired bioactivation of clopidogrel, especially in patients having undergone percutaneous intervention, and in particular for stent thrombosis as the major cardiovascular event (Mega *et al.*, 2010).

Computer-inferred haplotypes are generally avoided, although some predicted alleles have been assigned and included on the CYP-allele website. Alleles published in the early beginning of the CYP-allele era do not fully comply with the current standard.

Submission of new alleles is achieved by contacting the Webmaster of the CYP-allele website, whereby the data characterizing the allele is reviewed for potential allele name designation (see inclusion criteria at <http://www.cypalleles.ki.se/criteria.htm>). All information is kept confidential and a manuscript in preparation can often serve as a good basis for review and discussion between the author and the Webmaster. Designation of allele names outside the CYP-allele nomenclature committee is not advised, due to the apparent risk of confusion in the literature. Submissions with respect to additional functional in vivo and in vitro characterization of alleles listed are also highly relevant, and aids in keeping the CYP-allele website up to date,

Deleted: a

Deleted: Inclusion

Deleted: .

Deleted: known

Deleted: with respect to associated effects observed for example *in vitro* and *in vivo*

4. The Human Arylamine N-Acetyltransferase (NAT) Gene Nomenclature

Committee

Arylamine N-acetyltransferases (NATs) conjugate drugs and xenobiotics containing arylamine structures and these enzymes were early described in terms of pharmacogenetic variation, resulting in slow and rapid acetylation phenotypes that may impact the risk of drug-induced toxicity (Sim et al., 2008). There are two active genes in the *NAT* locus, *NAT1* and *NAT2* and in 1995, a nomenclature system was published (Vatsis et al., 1995). However, a systematic procedure for naming new alleles was not in use at once and resulted in problems such as dual assignments. From 1998, further discussions resulted in the formation of the Human Arylamine N-Acetyltransferase (*NAT*) Gene Nomenclature Committee with the aim to reach a nomenclature consensus for *NAT* alleles (Hein et al., 2008; Hein et al., 2000). A website was launched for maintenance of the renewed system and to provide an update on *NAT* alleles (<http://n-acetyltransferasenomenclature.louisville.edu/>). New allele submissions are sent to the *NAT* Nomenclature Committee for review, and the *NAT* allele website covers alleles of the *NAT1* and *NAT2* genes. The *NAT* allele nomenclature follows the star system (e.g. *NAT*5*) with special focus on functionally relevant sequence variations, whereby sub-alleles defined by an extra letter may contain additional variants that are functionally not relevant compared to the SNP causing the major effect. For example, alleles of the *NAT2*5* series (*NAT2*5A-P*) all contain a SNP causing the I114T amino acid substitution that in turn causes slow acetylation, which is associated with an increased risk of drug-induced liver injury in isoniazid-treated tuberculosis patients (Daly, 2010).

Deleted: for these

Deleted: immediately

Deleted: of reaching

Deleted: with

Deleted: causative

Deleted: secondary to the causative SNP (e.g. *NAT2*5B*)

The NAT website [also](#) contains *NAT* allele information in the pdf format reporting nucleotide changes, amino acid changes, phenotypes (slow or rapid acetylation), NCBI dbSNP numbers and relevant references, [as well as](#) information on allele inclusion criteria for designation of new alleles.

Deleted: . The *NAT* nomenclature pdf files also contain

As opposed to other nomenclature systems, the “wild type” allele is not designated *NAT*1* due to historical reasons. Thus, the reference allele for *NAT* nomenclature is *NAT*4* for both [NAT](#) genes. As a result of this, *NAT* alleles with numbers between *1 and *3 are excluded from the nomenclature system. At present, 29 alleles are presented for *NAT1* and 21 for *NAT2*.

Deleted: and the risk of confusion with renaming of alleles.

Deleted: *NAT1*

Deleted: mother

5. The Pharmacogenetics of Membrane Transporters (PMT) database

The Pharmacogenetics of Membrane Transporters (PMT) database

(<http://pharmacogenetics.ucsf.edu/>) (Kroetz et al., 2010) is part of the PharmGKB (see section 2) with as focus on drug response in relation to genetic variation in membrane transporters.

Since membrane transporters are involved in the transport of drugs, genetic variation affecting transport function may impact drug therapy in terms of both response and ADRs. The PMT project was initiated in 2000 and focuses on the two superfamilies of solute carrier (SLC) transporters ([n= 366](#)) and ATP binding cassette (ABC) transporters ([n=46](#)). The database covers coding and non-coding variations as well as both SNPs and [indels](#), but does not provide allele or haplotype information. The list of transporter proteins [contains](#) the OMIM phenotype, nucleotide and protein reference sequences, Entrez hits, and number of isoforms (splice variants) for every gene. The whole list is composed of 430 genes, although not all are [transporter](#) proteins since for some unknown reason [genes such as](#) monoamine oxidase A

Deleted: on

Deleted: its work

Deleted: .

Deleted: in/del's

Deleted: with

Deleted: PMT database main

Deleted: provide with

Deleted: transporters

Deleted: e.g.

(MAOA), pregnane X receptor (PXR) and POR are [also](#) included in this list. Most genes are extensively presented [on their own web page](#), containing information such as main tissue expression sites, substrates, gene structure, and genetic variation. The genetic variants are reported by their genomic, transcript and coding position, together with amino acid effects, array availability, PubMed and dbSNP links, and the frequency observed in different populations. Also, a secondary structure transmembrane prediction shows the position of amino acid substitutions, insertions or deletions within the transporter protein. Additionally, the PMT database provides a search tool to access gene expression data with respect to a single gene in multiple tissues (Gene Plot), selected genes in selected tissues (Mixed Plot) and multiple genes in a single tissue (Tissue Plot).

Deleted: The PMT database's POR web page contains additional apparent non-coding variants, but do not provide any additional causative variants, as compared to the official *POR* allele nomenclature present at the *CYP*-allele website, neither does it provide with allele or haplotype information. Of the 430 genes presented, 366 encode SLC transporters and 46 encode ABC transporters.

Deleted: with its

Deleted: with

6. Transporter database (TP-search)

On the Transporter Database TP-search [website \(http://www.tp-search.jp/\)](http://www.tp-search.jp/) (Ozawa et al., 2004), transporters are summarized [under](#) the respective species and the human section contains 33 transporter proteins together with information such as the driving force used (e.g. ATP, [Na⁺](#)), the presence of SNPs, protein length, direction of transport and association with disease. In addition, searches can be performed by transporter name, compound name, drug-drug interaction, pathophysiology, genetic polymorphisms, and tissue name. Additional web pages include gender differences, animal models, and genetic diseases. The [Transporter Name](#) search includes the choice of human, mouse and/or rat species, and substrate, inhibitor and/or inducer substances. The [Transporter Name](#) search outcome includes parameters associated [with](#) the requested transporter, [i.e.](#) substance name and chemical structure, interaction type,

Deleted: (

Deleted:)

Deleted: for

Deleted: NA+, P+),

Deleted: to this

Deleted: knock-out and mutant

Deleted: transporter name

Deleted: transporter name

Deleted: to

Deleted: ;

K_m and K_i values, experimental setup used for producing the data and references to publications. Similar searches can be done with the [Compound Name](#), where a list of transporters interacting with the substance is provided. The [Drug-Drug Interaction](#) search involves *in vivo* data, and displays drug-drug interactions involving the searched compound, together with its effects on parameters such as clearance, C_{max} and area under plasma concentration curve (AUC). Also, the interaction site where the effects are believed to [occur](#) is presented. The [Tissue Distribution](#) search provides schematic figures of [the transporters](#) present in intestine, kidney [or](#) liver. In addition, a tissue distribution table lists the tissue and localization of the transporters, together with the experimental method used and the corresponding references. The [Genetic Polymorphism](#) search of functionality yields *in vivo* information on the respective transporter SNPs (but not alleles or haplotypes) with respect to nucleotide changes, protein changes, drug involved, type of drug administration, number and ethnicity of the study sample, *in vivo* parameters (such as AUC or plasma concentration), and the relationship between the different genotypes. A similar [Genetic](#) search with respect to frequency yields the distribution of the polymorphisms in different populations. A reference sequence with respect to the nucleotide change is however not provided and the TP-search database has not been updated since [2007](#).

Formatted: Subscript

Formatted: Subscript

Deleted: compound name

Deleted: drug-drug interaction

Formatted: Subscript

Deleted: arise

Deleted: tissue distribution

Deleted: with

Deleted: which transporter is

Deleted: and

Deleted: , and in which direction they are working.

Deleted: genetic polymorphism

Deleted: ,

Deleted: genetic

Deleted: it appears that

Deleted: 2006

[7. UGT Alleles Nomenclature Page](#)

[UGT enzymes \(Phase II metabolism\) link glycosyl groups to endogenous and exogenous substrates including drugs, to yield more hydrophilic products that can be more easily excreted from the body \(Mackenzie *et al.*, 2005\). The UDP-glucuronosyltransferase \(UGT\) Allele](#)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

[Nomenclature Page \(http://www.pharmacogenomics.pha.ulaval.ca/sgc/ugt_alleles\)](http://www.pharmacogenomics.pha.ulaval.ca/sgc/ugt_alleles) contains [UGT enzymes of subfamily UGT1A \(9 enzyme web pages\)](#) and [UGT2B \(5 active web pages\)](#). The [inclusion criteria are based on the CYP-allele nomenclature system, and thus alleles designated by the Webmaster contains an asterisk and a number where the *1 allele \(e.g. *UGT1A1*1*\) is considered the wild type allele. Similarly to CYP-alleles, the UGT Alleles Nomenclature Page focus on nucleotide changes that cause functional effects such as amino acid substitutions, splice defects and frameshifts, and thus alleles with additional variations not causing any effects receives a sub \(letter\) allele name. Every enzyme has a “Haplotype” page and a “SNP” page, where the former contains a list of all alleles and their changes whereas the latter contain a list of all SNPs together with their rs numbers and the reference sequence used. As an example, the \[UGT1A1 Haplotype web page\]\(#\) contains 113 different alleles that are presented with their allelic names, nucleotide changes, protein effects, exon numbers affected, consequences \(e.g. frameshift\), phenotypes with respect to causation of Crigler-Najjar syndrome or Gilbert’s syndrome, enzyme activity *in vitro* and *in vivo*, and the respective references. The rare Crigler-Najjar syndrome is caused by a complete inability to glucuronize bilirubin due to \[UGT1A1 mutations that consequently causes jaundice and potential brain injury, whereas reduced but not depleted UGT1A1-mediated glucuronidation of bilirubin lead to the less severe phenotype Gilbert’s syndrome. UGT1A1 is also involved in the metabolism of the anti-cancer drug irinotecan, since the active metabolite SN-38 is glucuronidated and inactivated by UGT1A1. Excessive levels of SN-38 can cause neutropenia, the risk of which is affected by \\[UGT1A1 genotype. The wild-type allele *UGT1A1*1* carries a 6-repeat TATA box, \\\(A\\\(TA\\\)₆TAA\\\) whereas *UGT1A1*36* has 5, *UGT1A1*28* has 7 and *UGT1A1*37* has 8 repeats. An increased number of\\]\\(#\\)\]\(#\)](#)

[TA repeats causes reduced UGT1A1 gene transcription and the common reduced activity](#)

[UGT1A1*28 allele associates with reduced clearance of SN-38 \(Schulz 2009 PMID: 19770637\).](#)

[Due to the increased risk of neutropenia in UGT1A1*28/*28 subjects, the Food and Drug](#)

[Administration in the US recommends a lower starting dose of irinotecan in such patients.](#)

Formatted: Level 1, Keep with next,
 Tabs: 113.4 pt, Left + 177.2 pt,
 Left + 191.4 pt, Left + 269.35 pt,
 Left + 361.5 pt, Left + 375.65 pt,

8. PharmaADME

PharmaADME (<http://www.pharmaadme.org/>) is a database initiated by academic and

Deleted: (

Deleted:)

industrial interests. The aim of this database is to compile lists of genes and genetic variation

affecting Absorption, Distribution, Metabolism and Excretion (ADME) of drugs. The

PharmaADME's main aim is to aid in selection of genotyping assays for product development

and clinical trials. PharmaADME provides three gene lists, the ADME Gene List, the Core Marker

Deleted: in

Deleted: with

Deleted: regarded as

Gene List and the Related Gene List. The ADME Gene Lists contain genes associated with drug

metabolism, either sorted under the more important Core List of 32 genes or under the

Extended ADME Gene List of 267 genes. The ADME Gene Lists contains the gene symbol, full

gene name and the protein class, but no other information and no links to information on which

the qualifications are based or which drugs the gene product is affecting. The Related Gene List

contains 74 genes that are related to, but not involved in, drug metabolism such as drug targets

and receptors. The Core Marker Gene List is composed of 184 genotypes in 32 genes, of which

11 are CYPs, 9 are transporters and 9 are transferases, which by PharmaADME have been

selected as most important. This list comprises the gene symbol, the primary name (e.g.

*CYP2D6*5*) and the NCBI dbSNP number.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

9. Conclusions

The area of pharmacogenomics is rapidly developing. It is anticipated that novel, cheaper and

more efficient high throughput techniques for genomic analyses as well as the discovery of

interindividual variation in epigenomic control of gene expression and the identification of

additional types of snRNA molecules, will result in a rapid expansion of the area. Consequently,

the number of useful biomarkers for predicting drug response will rapidly increase in quantity

and quality. Thus, existing and future databases, which are developed in a user-friendly manner,

will provide relevant information on drugs and gene variants for physicians, scientists,

regulatory agencies and industry, and constitute very important tools for the possibility to

improve drug treatment and settle the basis for personalized medicine in society.

Deleted: 8

Deleted: in

Deleted: srRNA

Deleted: with

Deleted: the

Acknowledgements

The research at the laboratories of the authors is supported by grants from The Swedish

Research Council, The Swedish Brain Foundation, The Swedish cancer Fund and from NIH

GM61374.

REFERENCES

- Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB. 2010. Clinical assessment incorporating a personal genome. *Lancet* 375:1525-35.
- Bertz RJ, Granneman GR. 1997. Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions. *ClinPharmacokinet* 32:210-58.
- Coulet A, Shah NH, Garten Y, Musen M, Altman RB. 2010. Using text to build semantic networks for pharmacogenomics. *J Biomed Inform Aug 17*: [Epub ahead of print].
- [Daly AK. 2010. Drug-induced liver injury: past, present and future. *Pharmacogenomics* 11:607-11.](#) Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. 2009. Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient-episodes. *PLoS One* 4:e4439.
- [Davies EC, Green CF, Mottram DR, Rowe PH, Pirmohamed M. 2010. Emergency re-admissions to hospital due to adverse drug reactions within 1 year of the index admission. *Br J Clin Pharmacol* 70:749-55.](#) Eichelbaum M, Altman RB, Ratain M, Klein TE. 2009. New feature: pathways and important genes from PharmGKB. *Pharmacogenet Genomics* 19:403.
- Eichelbaum M, Ingelman-Sundberg M, Evans WE. 2006. Pharmacogenomics and individualized drug therapy. *Annu Rev Med* 57:119-37.
- [Evans WE, Relling MV. 1999. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286:487-91.](#)
- Garten Y, Altman RB. 2010. Teaching computers to read the pharmacogenomics literature ... so you don't have to. *Pharmacogenomics* 11:515-8.
- Garten Y, Altman RB. 2009. Pharmspresso: a text mining tool for extraction of pharmacogenomic concepts and relationships from full text. *BMC Bioinformatics* 10 Suppl 2:S6.
- Garten Y, Tatonetti NP, Altman RB. 2010. Improving the prediction of pharmacogenes using text-derived drug-gene relationships. *Pac Symp Biocomput*:305-14.
- Hansen NT, Brunak S, Altman RB. 2009. Generating genome-scale candidate gene lists for pharmacogenomics. *ClinPharmacolTher* 86:183-9.
- Hein DW, Boukouvala S, Grant DM, Minchin RF, Sim E. 2008. Changes in consensus arylamine N-acetyltransferase gene nomenclature. *Pharmacogenet Genomics* 18:367-8.
- Hein DW, Grant DM, Sim E. 2000. Update on consensus arylamine N-acetyltransferase gene nomenclature. *Pharmacogenetics* 10:291-2.
- Ingelman-Sundberg M. 2004. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends PharmacolSci* 25:193-200.
- Ingelman-Sundberg M, Oscarson M. 2002. Human CYP allele database: submission criteria procedures and objectives. *Methods Enzymol* 357:28-36.

Deleted: Evans DI. 1991. Thiamine responsive anaemia. *Br J Haematol* 78:140-1.¶

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Ingelman-Sundberg M, Sim SC, Gomez A, Rodriguez-Antona C. 2007. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoeigenetic and clinical aspects. *Pharmacol Ther* 116:496-526
- Kirchheiner J, Nickchen K, Bauer M, Wong ML, Licinio J, Roots I, Brockmoller J. 2004. Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. *Mol Psychiatry* 9:442-73.
- Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, Lee MT, Limdi NA, Page D, Roden DM, Wagner MJ, Caldwell MD, Johnson JA; International Warfarin Pharmacogenetics Consortium. 2009. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 360:753-64.
- Kroetz DL, Yee SW, Giacomini KM. 2010. The pharmacogenomics of membrane transporters project: research at the interface of genomics and transporter pharmacology. *ClinPharmacolTher* 87:109-16.
- Lazarou J, Pomeranz BH, Corey PN. 1998. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. [JAMA](#) 279:1200-5.
- Limdi NA, Wadelius M, Cavallari L, Eriksson N, Crawford DC, Lee MT, Chen CH, Motsinger-Reif A, Sagreiya H, Liu N, Wu AH, Gage BF, Jorgensen A, Pirmohamed M, Shin JG, Suarez-Kurtz G, Kimmel SE, Johnson JA, Klein TE, Wagner MJ; International Warfarin Pharmacogenetics Consortium. 2010. Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. *Blood* 115:3827-34.
- [Mackenzie PI, Bock KW, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, Miners JO, Owens IS, Nebert DW. 2005. Nomenclature update for the mammalian UDP glycosyltransferase \(UGT\) gene superfamily. *Pharmacogenet Genomics* 15:677-85.](#)
[Mega JL, Simon T, Collet JP, Anderson JL, Antman EM, Bliden K, Cannon CP, Danchin N, Giusti B, Gurbel P, Horne BD, Hulot JS, Kastrati A, Montalescot G, Neumann FJ, Shen L, Sibbing D, Steg PG, Trenk D, Wiviott SD, Sabatine MS. 2010. Reduced-function CYP2C19 genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. *JAMA* 304:1821-30.](#)
- Mi H, Thomas PD, Ring HZ, Jiang R, Sangkuhl K, Klein TE, Altman RB. 2010. PharmGKB summary: dopamine receptor D2. *Pharmacogenet Genomics* Aug 23: [Epub ahead of print].
- Moore TJ, Cohen MR, Furberg CD. 2007. Serious adverse drug events reported to the Food and Drug Administration, 1998-2005. *Arch Intern Med* 167: 1752-9.
- Oscarson M, Ingelman-Sundberg M. 2002. CYPalleles: a web page for nomenclature of human cytochrome P450 alleles. *Drug Metab Pharmacokinet* 17:491-5.
- Ozawa N, Shimizu T, Morita R, Yokono Y, Ochiai T, Munekata K, Ohashi A, Aida Y, Hama Y, Taki K, Maeda K, Kusuhara H, Sugiyama Y. 2004. Transporter database, TP-Search: a web-accessible comprehensive database for research in pharmacokinetics of drugs. *Pharm Res* 21:2133-4.
- Paul E, End-Rodrigues T, Thylen P, Bergman U. 2008. [Adverse drug reactions a common cause of hospitalization of the elderly. A clinical retrospective study]. *Lakartidningen* 105: 2338-42.

Deleted: Jama

Formatted: Swedish (Sweden)

Formatted: English (U.S.)

- 1
2 Pirmohamed M, Park BK. 2001. Genetic susceptibility to adverse drug reactions. *Trends*
3 *PharmacolSci* 22:298-305.
- 4 Sangkuhl K, Shuldiner AR, Klein TE, Altman RB. 2010. Platelet aggregation pathway.
5 *Pharmacogenet Genomics* Oct 8: [Epub ahead of print].
- 6 Sim E, Lack N, Wang CJ, Long H, Westwood I, Fullam E, Kawamura A. 2008. Arylamine N-
7 acetyltransferases: structural and functional implications of polymorphisms.
8 *Toxicology* 254:170-83.
- 9 Sim SC, Ingelman-Sundberg M. 2010. The Human Cytochrome P450 (CYP) Allele Nomenclature
10 website: a peer-reviewed database of CYP variants and their associated effects.
11 *Hum Genomics* 4:278-81.
- 12 Sim SC, Ingelman-Sundberg M. 2006. The human cytochrome P450 Allele Nomenclature
13 Committee Web site: submission criteria, procedures, and objectives. *Methods*
14 *MolBiol* 320: 183-91.
- 15 Sim SC, Ingelman-Sundberg M. 2011. Pharmacogenetic biomarkers: new tools in current and
16 future drug therapy. *Trends Pharmacol Sci*, in press.
- 17 Sim SC, Miller WL, Zhong XB, Arlt W, Ogata T, Ding X, Wolf CR, Flück CE, Pandey AV, Henderson
18 CJ, Porter TD, Daly AK, Nebert DW, Ingelman-Sundberg M. 2009. Nomenclature for
19 alleles of the cytochrome P450 oxidoreductase gene. *Pharmacogenet Genomics*
20 19(7):565-6.
- 21 [Spear BB, Heath-Chiozzi M, Huff J. 2001. Clinical application of pharmacogenetics. *Trends Mol*](#)
22 [Med 7:201-4.](#) Tatonetti NP, Liu T, Altman RB. 2009. Predicting drug side-effects by
23 chemical systems biology. *Genome Biol* 10:238.
- 24 Vatsis KP, Weber WW, Bell DA, Dupret JM, Evans DA, Grant DM, Hein DW, Lin HJ, Meyer UA,
25 Relling MV. 1995. Nomenclature for N-acetyltransferases. *Pharmacogenetics* 5:1-
26 17.
- 27 Weinshilboum R. 2003. Inheritance and drug response. *N Engl J Med* 348:529-37.
- 28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Formatted: English (U.S.)

Table 1. Pharmacogenetic databases.

<u>Database</u>	<u>Organisation</u>	<u>Activities</u>	<u>Web address</u>	<u>Content</u>
PharmGKB	The Pharmacogenomics Knowledgebase	Curation of pharmacogenetic effects, creation of gene and drug summaries, data sharing consortia	http://www.pharmgkb.org/	Genotype and phenotype of ADME genes and drug target genes
CYP-allele database	The Human Cytochrome P450 (CYP) Allele Nomenclature Website	Reviewing and designation of CYP alleles	http://www.cypalleles.ki.se/	CYP nomenclature and CYP allele characteristics
NAT-allele database	The Human Arylamine N-Acetyltransferase (NAT) Gene Nomenclature Committee	Reviewing and designation of NAT alleles	http://n-acetyltransferasenomenclature.louisville.edu/	NAT nomenclature and NAT allele characteristics
PMT database	The Pharmacogenetics of Membrane Transporters (PMT) database (part of PharmGKB)	Compilation of sequence variation in transporter genes (no alleles or haplotypes)	http://pharmacogenetics.ucsf.edu/	Genetic variation in SLC and ABCC transporters
TP-search database	The Transporter database	Summaries of transporters and their genetic variations (no alleles or haplotypes)	http://www.tp-search.jp/	Genotype and phenotype of SLC and ABCC transporters
UGT-allele database	The UDP-glucuronosyltransferase (UGT) Allele Nomenclature Page	Reviewing and designation of UGT alleles	http://www.pharmacogenomics.pha.ulaval.ca/sgc/ugt_alleles	UGT nomenclature and UGT allele characteristics
PharmaADME	Pharmacogenetics of Absorption, Distribution, Metabolism and Excretion genes	Compilation of ADME gene lists	http://www.pharmaadme.org/	Polymorphic ADME genes ranked in order of significance for causing variability in drug response

Formatted: English (U.S.)

Databases in the area of Pharmacogenetics

Sarah C Sim^{1/*}, Russ B Altman^{2/} and Magnus Ingelman-Sundberg^{1/}

^{1/} Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska

Institutet, SE-171 77 Stockholm, Sweden; ^{2/}Department of Bioengineering, Stanford University,
Stanford, California 94305, USA

*Correspondence to Sarah C Sim (Email: Sarah.Sim@ki.se, Phone +46 8 524 862 39, Fax +46 8
33 73 27)

ABSTRACT

In the area of pharmacogenetics and personalized health care it is obvious that databases providing important information of the occurrence and consequences of variant genes encoding drug metabolizing enzymes, drug transporters, drug targets and other proteins of importance for drug response or toxicity, are of critical value for scientists, physicians and industry. The primary outcome of the pharmacogenomic field is the identification of biomarkers that can predict drug toxicity and drug response, thereby individualizing and improving drug treatment of patients. The drug in question and the polymorphic gene exerting the impact are the main issues to be searched for in the databases. Here, we review the databases that provide useful information in this respect, of benefit for the development of the pharmacogenomic field.

Key words: Pharmacogenetics, database, personalized medicine, adverse drug reaction, cytochrome P450, transporter, polymorphism

1. Introduction

There is a rapid development in genetic analyses and functional genomics covering next generation sequencing methods and versatile arrays that rapidly analyze the most important mutations and gene copy number variations in the genome. Based on this progress, it becomes increasingly evident that much but not all of the drug therapy can be individualized based on genome sequence variation. These efforts can result in pharmacogenomic drug labels that advise the physician to use a genetic test prior to drug prescription. This test can be mandatory, recommended or for information purposes only (see (Sim and Ingelman-Sundberg, 2011) for a review). The genetic variation of interest mainly concerns genes encoding drug metabolizing enzymes, drug transporters, drug targets and human leukocyte antigens (HLA). Only 30-60 % of patients respond to common drug therapy using e.g. antidepressants, statins, and antipsychotics (Spear *et al.*, 2001). Furthermore, adverse drug reactions (ADRs) are an increasing problem (Moore *et al.*, 2007) and apparently cost society as much as the drug therapy itself (cf. Eichelbaum *et al.*, 2006). In addition, meta-analyses by Lazarou *et al.* (1998) revealed consequences of more than 100,000 deaths annually caused by ADRs in the US, and Davies *et al.* (2009) have documented an increased length of hospitalization for more than 25% of patients experiencing ADRs in the UK. Also, ADRs are shown to occur in 16 % of hospitalized patients, frequently with severe consequences (Davies *et al.*, 2009), and ADRs have in several studies been shown to be responsible for about 7 % of hospitalizations in the average population (Lazarou *et al.*, 1998; Davies *et al.*, 2009), 20 % of readmissions to the hospital (Davies *et al.*, 2010) and more than 30 % among elderly (Paul *et al.*, 2008). These data

1
2
3 emphasize the need of updated databases that can provide fruitful guidance to both scientists,
4
5
6 physicians, regulatory agencies and industries to cope with this major problem in human health.
7
8
9

10 **2. The Pharmacogenomics Knowledge Base (PharmGKB)**

11
12 The Pharmacogenomics Knowledgebase (PharmGKB, <http://www.pharmgkb.org/>) was
13
14 established in 2000 and strives to provide information on how human genetic variation impacts
15
16 drug response. Stable government-funded efforts such as Genbank, Gene Expression Omnibus,
17
18 and dbGAP from the National Center for Biomedical Information (NCBI) at the National
19
20 Institutes of Health (NIH) in the US are the preferred databases for archiving data, whereas
21
22 curation of genes (or groups of related homologous genes) and their variations are best handled
23
24 by locus-specific databases (LSDBs) such as the Human Cytochrome P450 (CYP) Allele
25
26 Nomenclature Website (see section 3), and others focusing on transporters, receptors, and
27
28 kinases. In contrast, PharmGKB focuses on four main activities: (1) curation of published
29
30 literature in respect to relationships between genes, their variations, and drug response
31
32 phenotypes, (2) creation of "Very Important Pharmacogene" (VIP gene) summaries that
33
34 highlight specific pharmacogenomic significance of genes that are very important for drug
35
36 response, (3) creation of pathway diagrams focusing on drug response (pharmacodynamics, PD)
37
38 and drug metabolism (pharmacokinetics, PK) pathways, and finally (4) assistance to organize
39
40 and lead data-sharing consortia within pharmacogenomics where multiple centers combine
41
42 data and expertise to answer questions that cannot be answered by any center alone (usually
43
44 because of statistical power and smaller data sets).
45
46
47
48
49
50
51
52
53
54

55 **2.1 Curation of published literature**

56
57
58
59
60

1
2
3 The core activity of PharmGKB curators is to survey the literature and capture reports of human
4
5 genetic variation that impacts drug response phenotypes. These phenotypes may occur at the
6
7 molecular (e.g. binding), cellular (e.g. expression), tissue (e.g. drug transformation) or
8
9 organismal level (e.g. disease, symptoms), and all annotations are encoded with respect to
10
11 these categories to assist in searches. The PharmGKB has a list of journals publishing
12
13 pharmacogenomics observations, and also broadly searches the literature on a weekly basis to
14
15 find associations between genes, their variations, drugs, and phenotypes. The PharmGKB uses
16
17 the Human Genome Nomenclature Committee recommendations for genes and variations. Also,
18
19 they use a combination of the RcNorm drug list with some modifications for drug, and Medical
20
21 Subject Headings (MeSH) and Unified Medical Language System (UMLS) to encode phenotypes.
22
23 The result of this curation activity is currently about 2500 annotations that associate particular
24
25 human variations (mostly SNPs, but also including insertions or deletions (indels), copy number
26
27 variations, structural changes, and haplotypes) with drug-related phenotypes. These
28
29 annotations were recently used to create a "pharmacogenomic annotation" of a complete
30
31 human genome sequence (Ashley et al., 2010). Interestingly, there were nearly 100 drugs for
32
33 which risks or benefits could be predicted for this individual. Many of them were "low
34
35 confidence" based on the level of published evidence, but several of them would have some
36
37 clinical utility. The application of the PharmGKB annotations to this genome was manual, and so
38
39 current efforts are aimed at representing the basic variant annotations and their potential
40
41 clinical consequences in a more structured manner that will allow more automated application
42
43 to individual genomes.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2.2 Writing VIP gene summaries

As the PharmGKB curators accumulate individual variant annotations, the genes may appear as either (1) repeatedly involved in drug response because of their role in general drug pharmacokinetics (e.g. cytochrome P450s, transporters, transferases) or (2) having an important role in the action of a single drug or class of drugs (e.g. pharmacodynamic drug targets). When the amount of information for these genes reaches a level that the curators find to be sufficiently large, they can designate these genes as VIP genes¹. The PharmGKB curators prioritize these genes for short written summary reviews of their role in drug response, and the key variations in these genes that have been shown to modulate drug response. These reviews are re-examined every two years or more frequently if the pace of new knowledge generation is rapid, and are written in a templated manner that includes names and alternate names, introductory information, key publications, key pathways, known drugs/substrates, associated phenotypes/diseases and summaries of key variants with their associated effects. There are currently 43 VIP genes on the PharmGKB website, and their summaries are published regularly in a cooperative venture with the *Pharmacogenetics & Genomics* journal in order to draw attention to them and to provide a citation for their usage (Eichelbaum et al., 2009; Mi et al., 2010).

2.3 Creating pathways of drug response and metabolism

¹Very Important Pharmacogene (VIP gene) denotes genes that have a documented role in modulating the response to one or more drugs

1
2
3 In addition to the VIP gene summaries, curators also create annotated graphical pathways
4 depicting the relationships of genes and drugs. These pathways are the most popular feature at
5 PharmGKB. Each pathway is the result of a PharmGKB curator gathering all information
6 available for the drug of interest (for example, by collecting the basic annotations and gene-
7 drug relationships discussed above), and assembling a pathway summary. These pathways are
8 often created collaboratively with experts working on these systems and they include a down-
9 loadable graphical depiction that is linked to the individual genes and small molecules depicted.
10 It also includes supporting data and references upon which the pathway is based. As with the
11 VIP genes, the pathways are also published regularly in the *Pharmacogenetics & Genomics*
12 journal (Eichelbaum et al., 2009; Sangkuhl et al., 2010). In addition to PharmGKB-generated
13 pathways, PharmGKB links to pathways from external resources, and they are made available to
14 augment the PharmGKB pathways.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

36 2.4 Data sharing consortia

37
38 A somewhat unexpected activity for PharmGKB has been the convening of data sharing
39 consortia. The mission of PharmGKB is to promote the creation and sharing of data about
40 pharmacogenomics, and as a trusted third party, PharmGKB is in a good position to bring
41 together groups of investigators interested in common scientific questions. After signing a
42 memorandum of understanding that outlines the responsibilities and privileges of joining the
43 consortium (and protects the rights of data generators to publish, as well as the rights of the
44 consortium members to see data and participate in its analysis), the PharmGKB accepts data
45 sets, brings them into a common format, and then participates in the statistical analysis to
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 answer one or more shared scientific questions. The first example of a PharmGKB consortium
4
5 was the International Warfarin Pharmacogenetics Consortium (IWPC) which brought together
6
7
8 21 groups internationally to create a globally-relevant warfarin dosing algorithm, as well as an
9
10 analysis of inter-ethnicity differences in warfarin dosing (Klein et al., 2009; Limdi et al., 2010).
11
12 Other consortia have subsequently been established addressing the discovery of new genetic
13
14 variations relevant to warfarin, the role of CYP2D6 variations in tamoxifen response, and the
15
16 role of CYP2D6 in selective serotonin reuptake inhibitor (SSRI) response. A list of consortia and
17
18 their members are available on the "Contributors" tab of PharmGKB.
19
20
21
22
23
24
25

26 2.5 The PharmGKB website and other activities

27
28 In addition to a global search capability, users of PharmGKB are able to search for particular
29
30 kinds of information in five more focused search boxes: genes, variants, pathways, drugs, and
31
32 diseases. These search boxes can accept any text, but will only return results of one of these 5
33
34 specific types. In addition, the home page provides the PharmGKB curators "Favorite Papers"
35
36 that are updated every two weeks, and are available as an archive over the last four years. The
37
38 home page also offers an automatically created pharmacogenomics news feed, access to
39
40 tutorial information, introductions to pharmacogenetics, and access to tools.
41
42
43
44
45
46
47

48 Current new efforts at PharmGKB focus on:

- 49 • The development of automated natural language parsing tools to extract annotation
50
51 information from the literature automatically (Coulet et al., 2010; Garten and Altman 2010;
52
53 Garten and Altman 2009).
54
55
56
57
58
59
60

- The addition of chemoinformatics, such as drug structure similarity metrics to help predicting new gene-drug interactions (Garten et al., 2010; Hansen et al., 2009).
- The creation of systems biology methods to look at drug response with a more network-based view of the genes involved in producing complex drug responses (Hansen et al., 2009; Tatonetti et al., 2009).

3. The Human Cytochrome P450 (CYP) Allele Nomenclature Website

Cytochrome P450 enzymes are responsible for 75-80 % of all phase I-dependent metabolism, resulting in more polar metabolites that can be further conjugated in Phase II metabolism, and for 65-70 % of the clearance of clinically used drugs (Bertz and Granneman 1997; Evans and Relling 1999). CYP enzymes are encoded by 57 *CYP* genes and divided into families, of which 1-3 (*CYP1-3*) are generally involved in the metabolism of exogenous substances including drugs, whereas members of other families mainly are involved in metabolism of endogenous compounds. Variation in *CYP* genes results in four different phenotypes classically defined as ultrarapid, extensive, intermediate and poor metabolizers.

Characterization of the genetic polymorphism of CYP enzymes provides a basis for personalized medicine composing drug dose adjustment and selection of drug type (Ingelman-Sundberg 2004; Ingelman-Sundberg et al., 2007; Kirchheiner et al., 2004; Pirmohamed and Park 2001; Weinshilboum 2003)

To unite and summarize efforts focusing on the characterization of CYP polymorphism, the Human Cytochrome P450 Allele (*CYP*-allele) Nomenclature Committee was established in 1999, and the efforts resulted in agreed *CYP* nomenclature guidelines and the launching of the *CYP*-

1
2
3 allele website (<http://www.cypalleles.ki.se/>). The main purpose of the *CYP*-allele website is the
4
5 management of new allele designations based on recognized nomenclature guidelines,
6
7 facilitation of rapid publication, as well as providing a readily available summary of alleles and
8
9 their associated effects (Ingelman-Sundberg and Oscarson 2002; Oscarson and Ingelman-
10
11 Sundberg 2002; Sim and Ingelman-Sundberg 2006; Sim and Ingelman-Sundberg 2010).
12
13
14

15
16 At present, 29 polymorphic CYPs are included on the *CYP*-allele website, with the addition in
17
18 2008 of NADPH cytochrome P450 reductase (POR) (Sim et al., 2009) due to its elementary
19
20 function as electron donor to CYP enzymes. Every gene has its own webpage, listing the
21
22 associated alleles and their consequences on the genetic, molecular, *in vitro* and *in vivo* levels,
23
24 all together with relevant references and links to the NCBI dbSNP database.
25
26
27

28
29 The nomenclature system designates numbers (such as *CYP2B6*4*) for alleles that have been
30
31 thoroughly sequenced and well characterized in terms of linkage of the functionally relevant
32
33 SNP with other polymorphic sites. Functionally relevant in this aspect is regarded as a variation
34
35 causing measurable consequences such as amino acid substitutions, translation terminations or
36
37 splice defects. At a minimum, the open reading frame must have been sequenced. All sequence
38
39 variations are numbered based on A in the translation initiation codon as the reference
40
41 nucleotide (+1). Additional letter-designated sub-alleles of the same number (such as
42
43 *CYP2B6*4B*) contain the same functional SNP but additional sequence variations not causing
44
45 functional effects. A combination of functionally relevant variants are designated a new allele
46
47 number due to the potential of combinatorial effects, unless one of the variant is deleterious
48
49 and thus defines the allele name. As an example, *CYP2C19*2* carries a SNP causing a splice
50
51 defect of the transcribed RNA that is detrimental for enzyme function. This splice defect
52
53
54
55
56
57
58
59
60

1
2
3 (c681G>A) exists in different combinations with other SNPs causing amino acid substitutions
4
5
6 and the combination alleles have thus received allele names *CYP2C19*2A-D*. The effect of
7
8 *CYP2C19* genotype on clopidogrel (Plavix[®]) antiplatelet treatment of cardiovascular disease has
9
10 attained much focus since *CYP2C19* is responsible for bioactivation of the drug. A recent meta-
11
12 analysis of 9 studies and almost 10,000 patients showed a significantly increased risk of adverse
13
14 cardiovascular events in carriers of defective (mainly *CYP2C19*2*) alleles due to impaired
15
16 bioactivation of clopidogrel, especially in patients having undergone percutaneous intervention,
17
18 and in particular for stent thrombosis as the major cardiovascular event (Mega *et al.*, 2010).
19
20
21 Computer-inferred haplotypes are generally avoided, although some predicted alleles have
22
23 been assigned and included on the *CYP*-allele website. Alleles published in the early beginning of
24
25 the *CYP*-allele era do not fully comply with the current standard.

26
27
28
29
30
31 Submission of new alleles is achieved by contacting the Webmaster of the *CYP*-allele
32
33 website, whereby the data characterizing the allele is reviewed for potential allele name
34
35 designation (see inclusion criteria at <http://www.cypalleles.ki.se/criteria.htm>). All information is
36
37 kept confidential and a manuscript in preparation can often serve as a good basis for review and
38
39 discussion between the author and the Webmaster. Designation of allele names outside the
40
41 *CYP*-allele nomenclature committee is not advised, due to the apparent risk of confusion in the
42
43 literature. Submissions with respect to additional functional *in vivo* and *in vitro* characterization
44
45 of alleles listed are also highly relevant, and aids in keeping the *CYP*-allele website up to date.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

4. The Human Arylamine N-Acetyltransferase (NAT) Gene Nomenclature

Committee

Arylamine N-acetyltransferases (NATs) conjugate drugs and xenobiotics containing arylamine structures and these enzymes were early described in terms of pharmacogenetic variation, resulting in slow and rapid acetylation phenotypes that may impact the risk of drug-induced toxicity (Sim et al., 2008). There are two active genes in the *NAT* locus, *NAT1* and *NAT2* and in 1995, a nomenclature system was published (Vatsis et al., 1995). However, a systematic procedure for naming new alleles was not in use at once and resulted in problems such as dual assignments. From 1998, further discussions resulted in the formation of the Human Arylamine N-Acetyltransferase (*NAT*) Gene Nomenclature Committee with the aim to reach a nomenclature consensus for *NAT* alleles (Hein et al., 2008; Hein et al., 2000). A website was launched for maintenance of the renewed system and to provide an update on *NAT* alleles (<http://n-acetyltransferasenomenclature.louisville.edu/>). New allele submissions are sent to the *NAT* Nomenclature Committee for review, and the *NAT* allele website covers alleles of the *NAT1* and *NAT2* genes. The *NAT* allele nomenclature follows the star system (e.g. *NAT*5*) with special focus on functionally relevant sequence variations, whereby sub-alleles defined by an extra letter may contain additional variants that are functionally not relevant compared to the SNP causing the major effect. For example, alleles of the *NAT2*5* series (*NAT2*5A-P*) all contain a SNP causing the I114T amino acid substitution that in turn causes slow acetylation, which is associated with an increased risk of drug-induced liver injury in isoniazid-treated tuberculosis patients (Daly, 2010).

1
2
3 The NAT website also contains *NAT* allele information in the pdf format reporting nucleotide
4 changes, amino acid changes, phenotypes (slow or rapid acetylation), NCBI dbSNP numbers and
5
6 relevant references, as well as information on allele inclusion criteria for designation of new
7
8 alleles.
9
10

11
12 As opposed to other nomenclature systems, the “wild type” allele is not designated *NAT*1* due
13 to historical reasons. Thus, the reference allele for *NAT* nomenclature is *NAT*4* for both *NAT*
14 genes. As a result of this, *NAT* alleles with numbers between **1* and **3* are excluded from the
15 nomenclature system. At present, 29 alleles are presented for *NAT1* and 21 for *NAT2*.
16
17
18
19
20
21
22
23
24
25

26 **5. The Pharmacogenetics of Membrane Transporters (PMT) database**

27
28 The Pharmacogenetics of Membrane Transporters (PMT) database

29
30 (<http://pharmacogenetics.ucsf.edu/>) (Kroetz et al., 2010) is part of the PharmGKB (see section
31
32 2) with as focus on drug response in relation to genetic variation in membrane transporters.
33
34

35 Since membrane transporters are involved in the transport of drugs, genetic variation affecting
36 transport function may impact drug therapy in terms of both response and ADRs. The PMT
37 project was initiated in 2000 and focuses on the two superfamilies of solute carrier (SLC)
38 transporters (n= 366) and ATP binding cassette (ABC) transporters (n=46). The database covers
39 coding and non-coding variations as well as both SNPs and indels, but does not provide allele or
40 haplotype information. The list of transporter proteins contains the OMIM phenotype,
41 nucleotide and protein reference sequences, Entrez hits, and number of isoforms (splice
42 variants) for every gene. The whole list is composed of 430 genes, although not all are
43 transporter proteins since for some unknown reason genes such as monoamine oxidase A
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 (MAOA), pregnane X receptor (PXR) and POR are also included in this list. Most genes are
4
5 extensively presented on their own web page, containing information such as main tissue
6
7 expression sites, substrates, gene structure, and genetic variation. The genetic variants are
8
9 reported by their genomic, transcript and coding position, together with amino acid effects,
10
11 array availability, PubMed and dbSNP links, and the frequency observed in different
12
13 populations. Also, a secondary structure transmembrane prediction shows the position of
14
15 amino acid substitutions, insertions or deletions within the transporter protein. Additionally, the
16
17 PMT database provides a search tool to access gene expression data with respect to a single
18
19 gene in multiple tissues (Gene Plot), selected genes in selected tissues (Mixed Plot) and multiple
20
21 genes in a single tissue (Tissue Plot).
22
23
24
25
26
27
28
29
30

31 **6. Transporter database (TP-search)**

32
33 On the Transporter Database TP-search website (<http://www.tp-search.jp/>) (Ozawa et al.,
34
35 2004), transporters are summarized under the respective species and the human section
36
37 contains 33 transporter proteins together with information such as the driving force used (e.g.
38
39 ATP, Na⁺), the presence of SNPs, protein length, direction of transport and association with
40
41 disease. In addition, searches can be performed by transporter name, compound name, drug-
42
43 drug interaction, pathophysiology, genetic polymorphisms, and tissue name. Additional web
44
45 pages include gender differences, animal models, and genetic diseases. The Transporter Name
46
47 search includes the choice of human, mouse and/or rat species, and substrate, inhibitor and/or
48
49 inducer substances. The Transporter Name search outcome includes parameters associated
50
51 with the requested transporter, i.e. substance name and chemical structure, interaction type,
52
53
54
55
56
57
58
59
60

1
2
3 K_m and K_i values, experimental setup used for producing the data and references to
4
5 publications. Similar searches can be done with the Compound Name, where a list of
6
7 transporters interacting with the substance is provided. The Drug-Drug Interaction search
8
9 involves *in vivo* data, and displays drug-drug interactions involving the searched compound,
10
11 together with its effects on parameters such as clearance, C_{max} and area under plasma
12
13 concentration curve (AUC). Also, the interaction site where the effects are believed to occur is
14
15 presented. The Tissue Distribution search provides schematic figures of the transporters present
16
17 in intestine, kidney or liver. In addition, a tissue distribution table lists the tissue and localization
18
19 of the transporters, together with the experimental method used and the corresponding
20
21 references. The Genetic Polymorphism search of functionality yields *in vivo* information on the
22
23 respective transporter SNPs (but not alleles or haplotypes) with respect to nucleotide changes,
24
25 protein changes, drug involved, type of drug administration, number and ethnicity of the study
26
27 sample, *in vivo* parameters (such as AUC or plasma concentration), and the relationship
28
29 between the different genotypes. A similar Genetic search with respect to frequency yields the
30
31 distribution of the polymorphisms in different populations. A reference sequence with respect
32
33 to the nucleotide change is however not provided and the TP-search database has not been
34
35 updated since 2007.
36
37
38
39
40
41
42
43
44
45
46
47

48 **7. UGT Alleles Nomenclature Page**

49
50
51 UGT enzymes (Phase II metabolism) link glycosyl groups to endogenous and exogenous
52
53 substrates including drugs, to yield more hydrophilic products that can be more easily excreted
54
55 from the body (Mackenzie *et al.*, 2005). The UDP-glucuronosyltransferase (UGT) Allele
56
57
58
59
60

1
2
3 Nomenclature Page (http://www.pharmacogenomics.pha.ulaval.ca/sgc/ugt_alleles) contains
4
5 UGT enzymes of subfamily UGT1A (9 enzyme web pages) and UGT2B (5 active web pages). The
6
7 inclusion criteria are based on the *CYP*-allele nomenclature system, and thus alleles designated
8
9 by the Webmaster contains an asterisk and a number where the *1 allele (e.g. *UGT1A1*1*) is
10
11 considered the wild type allele. Similarly to *CYP*-alleles, the UGT Alleles Nomenclature Page
12
13 focus on nucleotide changes that cause functional effects such as amino acid substitutions,
14
15 splice defects and frameshifts, and thus alleles with additional variations not causing any effects
16
17 receives a sub (letter) allele name. Every enzyme has a “Haplotype” page and a “SNP” page,
18
19 where the former contains a list of all alleles and their changes whereas the latter contain a list
20
21 of all SNPs together with their rs numbers and the reference sequence used. As an example, the
22
23 UGT1A1 Haplotype web page contains 113 different alleles that are presented with their allelic
24
25 names, nucleotide changes, protein effects, exon numbers affected, consequences (e.g.
26
27 frameshift), phenotypes with respect to causation of Crigler-Najjar syndrome or Gilbert’s
28
29 syndrome, enzyme activity *in vitro* and *in vivo*, and the respective references. The rare Crigler-
30
31 Najjar syndrome is caused by a complete inability to glucuronize bilirubin due to UGT1A1
32
33 mutations that consequently causes jaundice and potential brain injury, whereas reduced but
34
35 not depleted UGT1A1-mediated glucuronidation of bilirubin lead to the less severe phenotype
36
37 Gilbert’s syndrome. UGT1A1 is also involved in the metabolism of the anti-cancer drug
38
39 irinotecan, since the active metabolite SN-38 is glucuronidated and inactivated by UGT1A1.
40
41 Excessive levels of SN-38 can cause neutropenia, the risk of which is affected by *UGT1A1*
42
43 genotype. The wild-type allele *UGT1A1*1* carries a 6-repeat TATA box, (A(TA)₆TAA) whereas
44
45 *UGT1A1*36* has 5, *UGT1A1*28* has 7 and *UGT1A1*37* has 8 repeats. An increased number of
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 TA repeats causes reduced *UGT1A1* gene transcription and the common reduced activity
4
5
6 *UGT1A1*28* allele associates with reduced clearance of SN-38 (Schulz 2009 PMID: 19770637).
7
8
9 Due to the increased risk of neutropenia in *UGT1A1*28/*28* subjects, the Food and Drug
10
11 Administration in the US recommends a lower starting dose of irinotecan in such patients.
12
13
14

15 16 **8. PharmaADME**

17
18 PharmaADME (<http://www.pharmaadme.org/>) is a database initiated by academic and
19
20 industrial interests. The aim of this database is to compile lists of genes and genetic variation
21
22 affecting Absorption, Distribution, Metabolism and Excretion (ADME) of drugs. The
23
24 PharmaADME's main aim is to aid in selection of genotyping assays for product development
25
26 and clinical trials. PharmaADME provides three gene lists, the ADME Gene List, the Core Marker
27
28 Gene List and the Related Gene List. The ADME Gene Lists contain genes associated with drug
29
30 metabolism, either sorted under the more important Core List of 32 genes or under the
31
32 Extended ADME Gene List of 267 genes. The ADME Gene Lists contains the gene symbol, full
33
34 gene name and the protein class, but no other information and no links to information on which
35
36 the qualifications are based or which drugs the gene product is affecting. The Related Gene List
37
38 contains 74 genes that are related to, but not involved in, drug metabolism such as drug targets
39
40 and receptors. The Core Marker Gene List is composed of 184 genotypes in 32 genes, of which
41
42 11 are CYPs, 9 are transporters and 9 are transferases, which by PharmaADME have been
43
44 selected as most important. This list comprises the gene symbol, the primary name (e.g.
45
46 *CYP2D6*5*) and the NCBI dbSNP number.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

9. Conclusions

The area of pharmacogenomics is rapidly developing. It is anticipated that novel, cheaper and more efficient high throughput techniques for genomic analyses as well as the discovery of interindividual variation in epigenomic control of gene expression and the identification of additional types of snRNA molecules, will result in a rapid expansion of the area. Consequently, the number of useful biomarkers for predicting drug response will rapidly increase in quantity and quality. Thus, existing and future databases, which are developed in a user-friendly manner, will provide relevant information on drugs and gene variants for physicians, scientists, regulatory agencies and industry, and constitute very important tools for the possibility to improve drug treatment and settle the basis for personalized medicine in society.

Acknowledgements

The research at the laboratories of the authors is supported by grants from The Swedish Research Council, The Swedish Brain Foundation, The Swedish cancer Fund and from NIH GM61374.

REFERENCES

- Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB. 2010. Clinical assessment incorporating a personal genome. *Lancet* 375:1525-35.
- Bertz RJ, Granneman GR. 1997. Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin Pharmacokinet* 32:210-58.
- Coulet A, Shah NH, Garten Y, Musen M, Altman RB. 2010. Using text to build semantic networks for pharmacogenomics. *J Biomed Inform* Aug 17: [Epub ahead of print].
- Daly AK. 2010. Drug-induced liver injury: past, present and future. *Pharmacogenomics* 11:607-11.
- Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. 2009. Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient-episodes. *PLoS One* 4:e4439.
- Davies EC, Green CF, Mottram DR, Rowe PH, Pirmohamed M. 2010. Emergency re-admissions to hospital due to adverse drug reactions within 1 year of the index admission. *Br J Clin Pharmacol* 70:749-55.
- Eichelbaum M, Altman RB, Ratain M, Klein TE. 2009. New feature: pathways and important genes from PharmGKB. *Pharmacogenet Genomics* 19:403.
- Eichelbaum M, Ingelman-Sundberg M, Evans WE. 2006. Pharmacogenomics and individualized drug therapy. *Annu Rev Med* 57:119-37.
- Evans WE, Relling MV. 1999. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286:487-91.
- Garten Y, Altman RB. 2010. Teaching computers to read the pharmacogenomics literature ... so you don't have to. *Pharmacogenomics* 11:515-8.
- Garten Y, Altman RB. 2009. Pharmspresso: a text mining tool for extraction of pharmacogenomic concepts and relationships from full text. *BMC Bioinformatics* 10 Suppl 2:S6.
- Garten Y, Tatonetti NP, Altman RB. 2010. Improving the prediction of pharmacogenes using text-derived drug-gene relationships. *Pac Symp Biocomput*:305-14.
- Hansen NT, Brunak S, Altman RB. 2009. Generating genome-scale candidate gene lists for pharmacogenomics. *Clin Pharmacol Ther* 86:183-9.
- Hein DW, Boukouvala S, Grant DM, Minchin RF, Sim E. 2008. Changes in consensus arylamine N-acetyltransferase gene nomenclature. *Pharmacogenet Genomics* 18:367-8.
- Hein DW, Grant DM, Sim E. 2000. Update on consensus arylamine N-acetyltransferase gene nomenclature. *Pharmacogenetics* 10:291-2.
- Ingelman-Sundberg M. 2004. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol Sci* 25:193-200.
- Ingelman-Sundberg M, Oscarson M. 2002. Human CYP allele database: submission criteria procedures and objectives. *Methods Enzymol* 357:28-36.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Ingelman-Sundberg M, Sim SC, Gomez A, Rodriguez-Antona C. 2007. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacol Ther* 116:496-526
- Kirchheiner J, Nickchen K, Bauer M, Wong ML, Licinio J, Roots I, Brockmoller J. 2004. Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. *Mol Psychiatry* 9:442-73.
- Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, Lee MT, Limdi NA, Page D, Roden DM, Wagner MJ, Caldwell MD, Johnson JA; International Warfarin Pharmacogenetics Consortium. 2009. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 360:753-64.
- Kroetz DL, Yee SW, Giacomini KM. 2010. The pharmacogenomics of membrane transporters project: research at the interface of genomics and transporter pharmacology. *ClinPharmacolTher* 87:109-16.
- Lazarou J, Pomeranz BH, Corey PN. 1998. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 279:1200-5.
- Limdi NA, Wadelius M, Cavallari L, Eriksson N, Crawford DC, Lee MT, Chen CH, Motsinger-Reif A, Sagreiya H, Liu N, Wu AH, Gage BF, Jorgensen A, Pirmohamed M, Shin JG, Suarez-Kurtz G, Kimmel SE, Johnson JA, Klein TE, Wagner MJ; International Warfarin Pharmacogenetics Consortium. 2010. Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. *Blood* 115:3827-34.
- Mackenzie PI, Bock KW, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, Miners JO, Owens IS, Nebert DW. 2005. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genomics* 15:677-85.
- Mega JL, Simon T, Collet JP, Anderson JL, Antman EM, Bliden K, Cannon CP, Danchin N, Giusti B, Gurbel P, Horne BD, Hulot JS, Kastrati A, Montalescot G, Neumann FJ, Shen L, Sibbing D, Steg PG, Trenk D, Wiviott SD, Sabatine MS. 2010. Reduced-function CYP2C19 genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. *JAMA* 304:1821-30.
- Mi H, Thomas PD, Ring HZ, Jiang R, Sangkuhl K, Klein TE, Altman RB. 2010. PharmGKB summary: dopamine receptor D2. *Pharmacogenet Genomics* Aug 23: [Epub ahead of print].
- Moore T J, Cohen MR, Furberg CD. 2007. Serious adverse drug events reported to the Food and Drug Administration, 1998-2005. *Arch Intern Med* 167: 1752-9.
- Oscarson M, Ingelman-Sundberg M. 2002. CYPalleles: a web page for nomenclature of human cytochrome P450 alleles. *Drug MetabPharmacokinet* 17:491-5.
- Ozawa N, Shimizu T, Morita R, Yokono Y, Ochiai T, Munesada K, Ohashi A, Aida Y, Hama Y, Taki K, Maeda K, Kusuhara H, Sugiyama Y. 2004. Transporter database, TP-Search: a web-accessible comprehensive database for research in pharmacokinetics of drugs. *Pharm Res* 21:2133-4.
- Paul E, End-Rodrigues T, Thylen P, Bergman U. 2008. [Adverse drug reactions a common cause of hospitalization of the elderly. A clinical retrospective study]. *Lakartidningen* 105: 2338-42.

- 1
2
3 Pirmohamed M, Park BK. 2001. Genetic susceptibility to adverse drug reactions. *Trends*
4 *PharmacolSci* 22:298-305.
5
6 Sangkuhl K, Shuldiner AR, Klein TE, Altman RB. 2010. Platelet aggregation pathway.
7 *Pharmacogenet Genomics* Oct 8: [Epub ahead of print].
8
9 Sim E, Lack N, Wang CJ, Long H, Westwood I, Fullam E, Kawamura A. 2008. Arylamine N-
10 acetyltransferases: structural and functional implications of polymorphisms.
11 *Toxicology* 254:170-83.
12 Sim SC, Ingelman-Sundberg M. 2010. The Human Cytochrome P450 (CYP) Allele Nomenclature
13 website: a peer-reviewed database of CYP variants and their associated effects.
14 *Hum Genomics* 4:278-81.
15
16 Sim SC, Ingelman-Sundberg M. 2006. The human cytochrome P450 Allele Nomenclature
17 Committee Web site: submission criteria, procedures, and objectives. *Methods*
18 *MolBiol* 320: 183-91.
19
20 Sim SC, Ingelman-Sundberg M. 2011. Pharmacogenetic biomarkers: new tools in current and
21 future drug therapy. *Trends Pharmacol Sci*, in press.
22
23 Sim SC, Miller WL, Zhong XB, Arlt W, Ogata T, Ding X, Wolf CR, Flück CE, Pandey AV, Henderson
24 CJ, Porter TD, Daly AK, Nebert DW, Ingelman-Sundberg M. 2009. Nomenclature for
25 alleles of the cytochrome P450 oxidoreductase gene. *Pharmacogenet Genomics*
26 19(7):565-6.
27
28 Spear BB, Heath-Chiozzi M, Huff J. 2001. Clinical application of pharmacogenetics. *Trends Mol*
29 *Med* 7:201-4. Tatonetti NP, Liu T, Altman RB. 2009. Predicting drug side-effects by
30 chemical systems biology. *Genome Biol* 10:238.
31
32 Vatsis KP, Weber WW, Bell DA, Dupret JM, Evans DA, Grant DM, Hein DW, Lin HJ, Meyer UA,
33 Relling MV. 1995. Nomenclature for N-acetyltransferases. *Pharmacogenetics* 5:1-
34 17.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Pharmacogenetic databases.

Database	Organisation	Activities	Web address	Content
PharmGKB	The Pharmacogenomics Knowledgebase	Curation of pharmacogenetic effects, creation of gene and drug summaries, data sharing consortia	http://www.pharmgkb.org/	Genotype and phenotype of ADME genes and drug target genes
CYP-allele database	The Human Cytochrome P450 (CYP) Allele Nomenclature Website	Reviewing and designation of CYP alleles	http://www.cypalleles.ki.se/	CYP nomenclature and CYP allele characteristics
NAT-allele database	The Human Arylamine N-Acetyltransferase (NAT) Gene Nomenclature Committee	Reviewing and designation of NAT alleles	http://n-acetyltransferasenomenclature.louisville.edu/	NAT nomenclature and NAT allele characteristics
PMT database	The Pharmacogenetics of Membrane Transporters (PMT) database (part of PharmGKB)	Compilation of sequence variation in transporter genes (no alleles or haplotypes)	http://pharmacogenetics.ucsf.edu/	Genetic variation in SLC and ABCC transporters
TP-search database	The Transporter database	Summaries of transporters and their genetic variations (no alleles or haplotypes)	http://www.tp-search.jp/	Genotype and phenotype of SLC and ABCC transporters
UGT-allele database	The UDP-glucuronosyltransferase (UGT) Allele Nomenclature Page	Reviewing and designation of UGT alleles	http://www.pharmacogenomics.pha.ulaval.ca/sgc/ugt_alleles	UGT nomenclature and UGT allele characteristics
PharmaADME	Pharmacogenetics of Absorption, Distribution, Metabolism and Excretion genes	Compilation of ADME gene lists	http://www.pharmaadme.org/	Polymorphic ADME genes ranked in order of significance for causing variability in drug response