



Stanford - South Africa

Biomedical Informatics Program



Building the PharmGKB: a web tool for pharmacogenomics

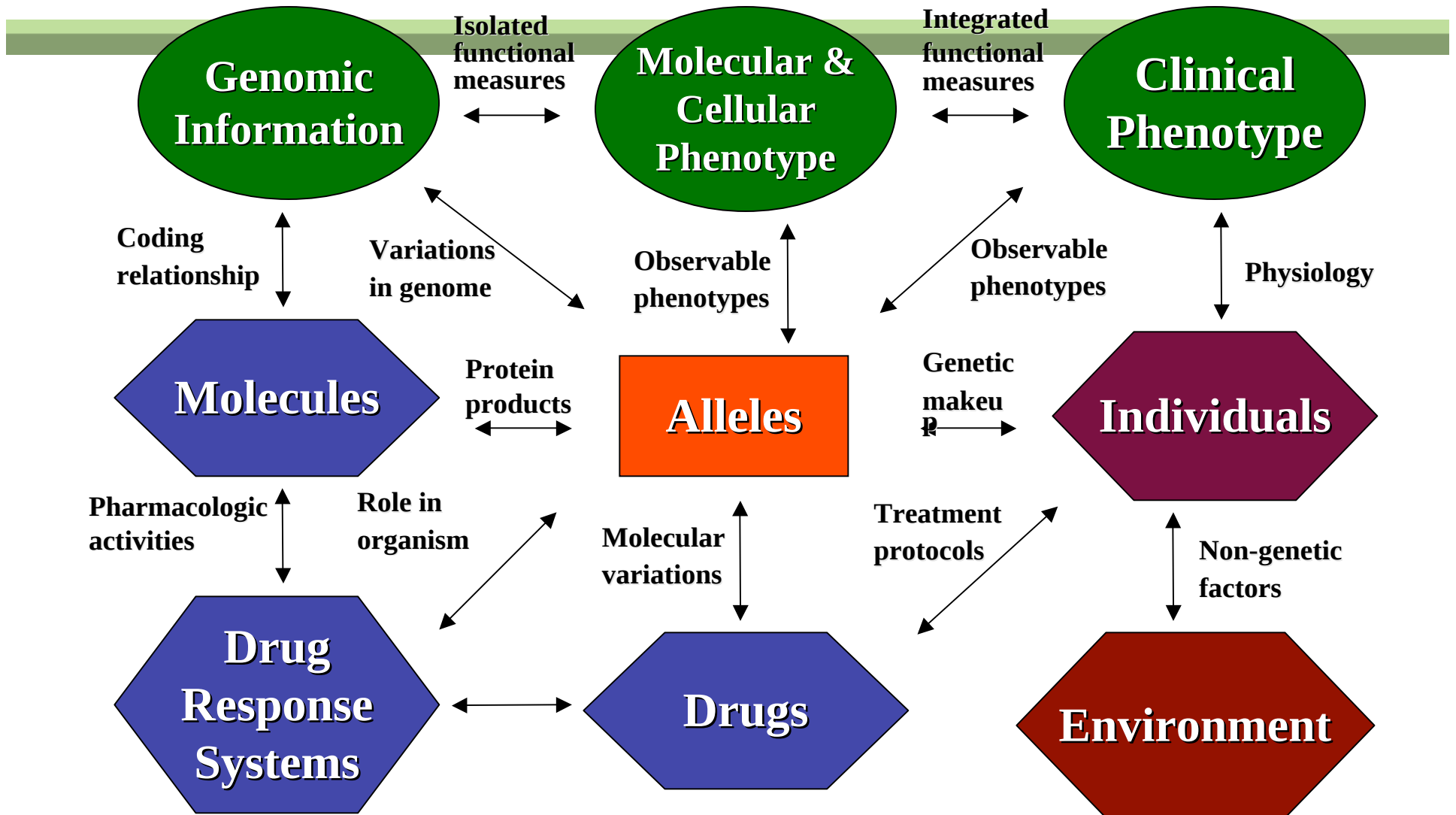
Russ B. Altman, MD, PhD

Professor of Genetics, Bioengineering &
Medicine

Stanford University



Complexity of Relationships in Pharmacogenetics



A public pharmacogenetics research network & database

- Establish a public database presence for pharmacogenetics (genotype-phenotype)
- Produce an available data set about genetic and phenotypic variability in response to medications.
- Concentrate on experimental systems with broad applicability for understanding variations in drug response.



<http://www.nigms.nih.gov/pharmacogenetics/>

Other Feeds ▾ CNN News Feed CNNSI Feed Slashdot Feed Science Magazine F... NY Times ▾ SJ Merc ▾ Nature Feeds ▾ Stanford Weather

NIGMS Home
What's New
Site Map / Index
Staff / Contact Us

Search NIGMS for: Go!

Text size: A A A

E-mail this link

RESEARCH FUNDING

Pharmacogenetics Research Network

PGRN Home News & Events Research Network Ethics & Communities Funding Opportunities Related Resources

PharmGKB Has New Features: Visual genotype and assayed regions are now available for genes. For example, see [SCN5A](#) on the [PharmGKB Web site](#). Click on the  icon (labelled "Genotype grid") on the top right for the visual genotype.


PGRN

The **Pharmacogenetics Research Network (PGRN)** is a nationwide collaboration of scientists studying the effect of genes on people's responses to a wide variety of medicines. See the [Goals Statement](#) of the NIH and the PGRN.

PharmGKB

The **Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB)** is an integrated knowledge base for pharmacogenetics linking phenotypes and genotypes.

News & Events

Meeting information, reports, news releases, and announcements about the Pharmacogenetics Research Network.

Research Network

Links to Pharmacogenetics

"One Size Doesn't Fit All"

Pharmacogenetics is the study of how genes affect the way people respond to medicines, including antidepressants, chemotherapy, drugs for asthma and heart disease, and many others. The ultimate goal of pharmacogenetics research is to help tailor medicines to people's unique genetic make-ups. This will make medicines safer and more effective for everyone.



PGRN II

Cardiovascular & Pulmonary

Statins & ACE Inhibitors

R. Krauss

Lawrence Berkeley National Lab

Antihypertensive response

J. Johnson

U. Florida

Cardiovascular risks

A. Shuldiner

U. Maryland

Arrhythmia Therapy

D. Roden

Vanderbilt U.

Asthma Treatment

S. Weiss

Harvard U.

PharmGKB

R. Altman

Stanford University

Cancer

Tamoxifen

D. Flockhart

Indiana University

Anti-Cancer Agents

M. Ratain

University of

Chicago

Drug Pathways, Metabolism and Transporters

Membrane Transporters

K. Giacomini

University of California, San Francisco

Drug Pathways Network

H. McLeod

Washington University

Phase II Drug Metabolizing Enzymes

R. Weinshilboum

Mayo Foundation

Nicotine Addiction

N. Benowitz

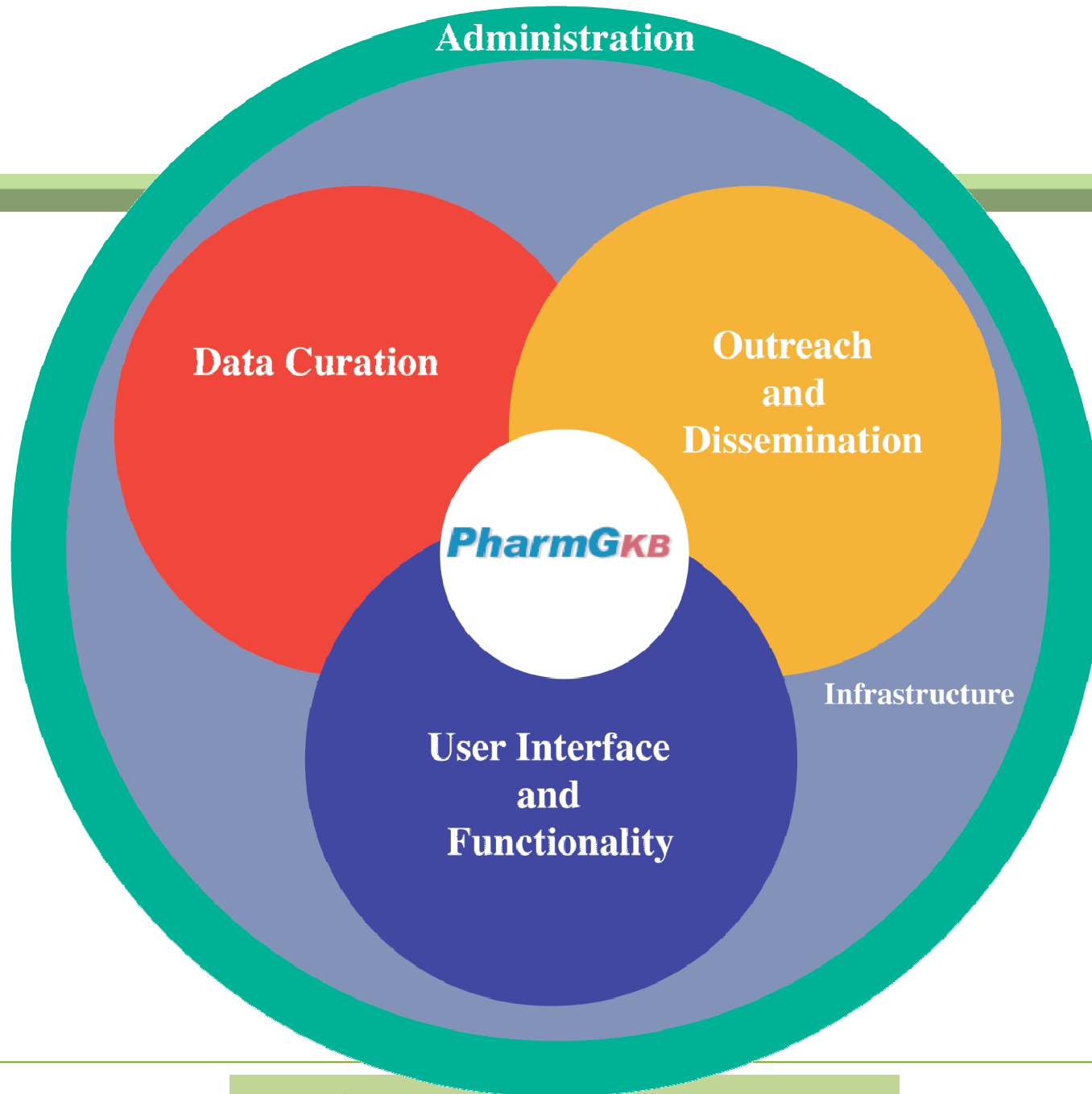
University of California, San Francisco



PharmGKB Goals

- **International data resource** of high quality data linking genotypes, phenotypes, pathways, and literature.
- **Curate data and develop exchange standards** for pharmacogenetics and pharmacogenomics
- Provide **analytic functionality** for users
- **Link** with complementary databases such as Medline, PDB (Protein Data Bank), dbSNP, and Genbank





PharmGKB
The Pharmacogenetics and Pharmacogenomics Knowledge Base

Search PharmGKB:

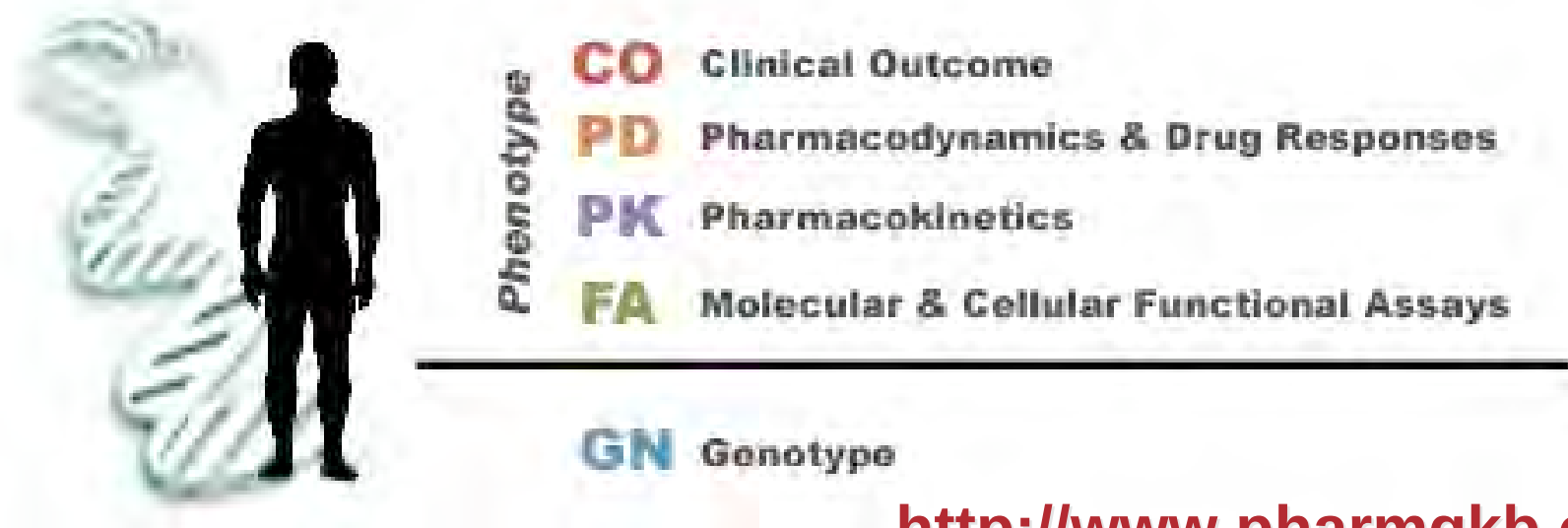
Home Search Submit Resources PGRN Contributors My PharmGKB sign in | help | feedback

Welcome | Overview | Events | Projects | Policies | Team | Register

The PharmGKB is an integrated resource about how variation in human genes leads to variation in our response to drugs. [more ...](#)

News
→ [PharmGKB Newsletter July 20 2007](#)

Genomic data, molecular and cellular phenotype data, and clinical phenotype data are accepted from the scientific community at large. These data are then organized and the relationships between genes and drugs are then categorized into the following categories:



<http://www.pharmgkb.org>



FA Molecular & Cellular Functional Assays

GN Genotype

→ [Other Data Submissions](#)
→ [PGRN Publications](#)
→ [FAQs on Publications](#)



NOS3

nitric oxide synthase 3 (endothelial cell)

Alternate Names: ECNOS; endothelial nitric oxidase synthase; endothelial nitric oxide synthase

Alternate Symbols: ECNOS; NOS III; eNOS

PharmGKB Accession ID:

PA254

[Download Data](#) 

Web Services

[Comparative Genomics](#)

[Add to Watchlist](#)

Cross-References

Entrez Gene ID:

[4846](#)

OMIM Accession:

[163729](#)

UCSC Genome Browser ID:

[NM_000603](#)

Ensembl ID:

[NOS3](#)

GDB ID:

GDB:209976

GO ID:

[GO:0004517](#)

GO ID:

PharmGKB Primary Data

Variant Positions:

[121 positions submitted](#)

Phenotype Data Sets:



[Folate, Homocysteine and Vitamin B12 in Young Healthy Individuals](#)

PD

Submitted by: [Alexander Steven Whitehead](#) involving [CBS](#), [MTHFR](#), [MTR](#), [MTRR](#), [NOS3](#), [TYMS](#), and [Hyperhomocysteinemia](#).



[Lipid measurements in tamoxifen study](#)

PD

GN

Submitted by: [David Flockhart, MD, PhD](#) involving [CYP2D6](#), [CYP3A5](#), [ESR1](#), [ESR2](#), [NOS3](#), [tamoxifen](#), and [Breast Neoplasms](#).



[Patient responses to tamoxifen](#)

PD

PK

GN

Submitted by: [David Flockhart, MD, PhD](#) involving [CYP2C9](#), [CYP2D6](#), [CYP3A5](#), [ESR1](#), [NOS3](#), [SULT1A1](#), [UGT1A6](#), [tamoxifen](#), and [Breast Neoplasms](#).

Pathways:

- [ACE-inhibitor Pathway](#)
- [VEGF Pathway](#)

Genotype data

Phenotype data

Pathway data

Literature annotation data

phenotype data available



genotype data available



literature annotations available

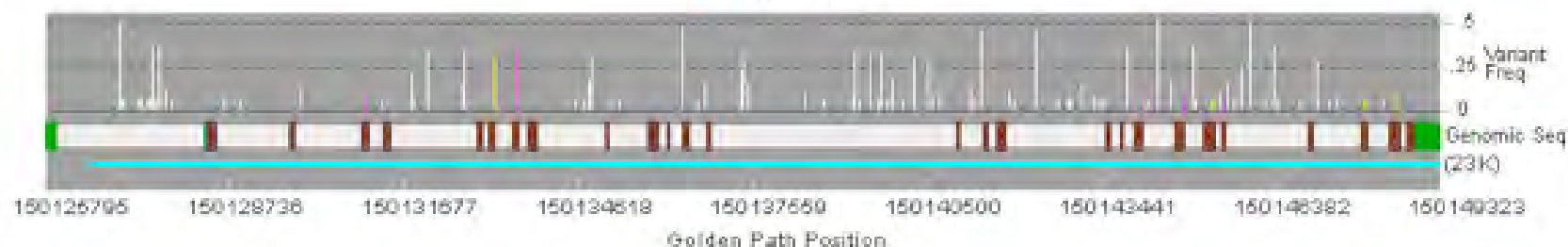
Literature Annotations

	Related Drugs	Relationship ?	Details
	 ace inhibitors	FA	
Φ	 acetylcholine	PD GN	
	 atorvastatin	FA	
	 estrogens	FA GN	
	 nitric oxide	PD GN	
	 nitroglycerin	PD GN	
Φ	 phenylephrine	CQ PD GN	
	 prostaglandins	PD GN	
	 rosuvastatin	PD	
	 sildenafil	FA	
Φ	 lammoxifen	CQ PD PK FA	

Variant Positions on NOS3

Web Services
[Comparative Genomics](#)

Gene | ☐ Intron ☒ Exon ☒ UTR ☒ Promoter ☒ Flanking
 Variant | ☐ Non-coding/Unknown ☒ Non-synonymous ☒ Synonymous
☒ Assayed Region ?



Click To: ☐ Move ☒ Magnify

Current Position: 150137559

Move:

Magnify: ?

Legend

- deletion

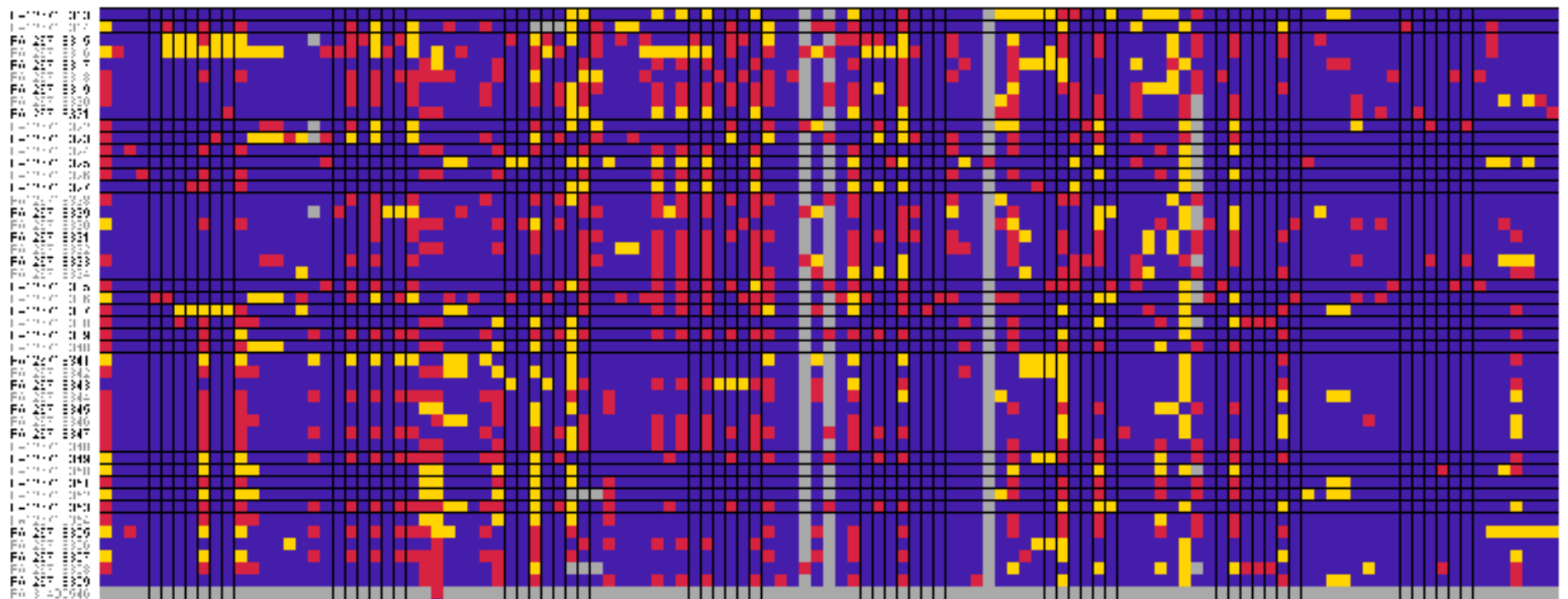
* feature derived from NCBI RefSeq[†]

Number of variant positions: 119

?

Golden Path Position	Variant	Strand	Feature	AA Translation	Frequency (%)	Sample Size	Phenotype	Submitted By	Submitted Position
chr7:150127045	T/A	plus	Intron*		52.13/47.87	94		PARC	450
chr7:150127074	A/T	plus	Intron*		98.94/1.06	94		PARC	479
chr7:150127138	G/T	plus	Intron*		97.87/2.13	94		PARC	543
chr7:150127384	C/T	plus	Intron*		98.94/1.06	94		PARC	789



[illegible]

3/22/07

Variants at chr7:150133759 on NOS3

Alleles: cccagatga G ccccccagaac [70.21%](#)
cccagatga T ccccccagaac [29.79%](#)

Submitted Position: 7164

Submitted By: [PARC](#)

Feature: Unknown

Assay(s) Used: [PCR Assay PA128452230*](#)

Legend

Frequency of Variants at Golden Path Position chr7:150133759 on NOS3

Submitted positions: GOFMOP 82
PARC 7164

Format:

Index

6,851 tccccaaaggt g
6,901 aggcaggaag a
6,951 gaggggtgcc t
7,001 atgaaggcag g
7,051 tcagccccag a
7,101 cggttggaac c
7,151 aggccccaga t
7,201 cttgagggtgc c
7,251 gtgggatgga g
7,301 ggtgctgata c
7,351 tgggcctgcg c
7,401 attggggggcc

	PARC Sample Set PA128275074		GOFMOP Sample Set PA134654021	
	Total	Not specified	Total	White (Caucasian)
Allele	94	94	696	696
G	66 (70.21%)	66 (70.21%)	465 (66.81%)	465 (66.81%)
T	28 (29.79%)	28 (29.79%)	231 (33.19%)	231 (33.19%)

[[Supporting Data](#)]

tggagttccc cgcagccccc ttcagtggct ggtacatgag

pack

full

hide

hide

hide

atics Program

Ensembl Genes

Acembly Genes

Twinscan

SGP Genes

Fgenes++ Genes

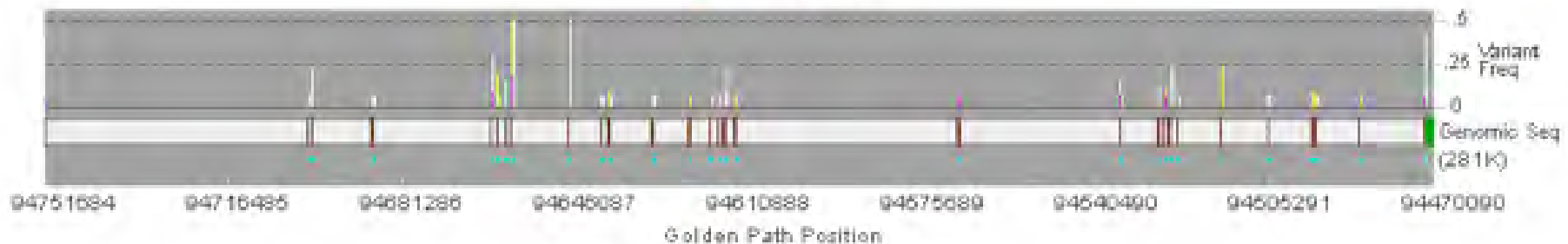


Variant Positions on ABCC4



Web Services
[Comparative Genomics](#)

Gene | ☐ Intron ☒ Exon ☒ UTR ☒ Promoter ☒ Flanking
 Variant | ☐ Non-coding/Unknown ☒ Non-synonymous ☒ Synonymous
☒ Assayed Region ?



Click To: ☐ Move ☒ Magnify

Current Position: 94610887

Move:

Magnify: ?

Legend

- deletion

* feature derived from NCBI RefSeq[†]

Number of variant positions: 99

?

Golden Path Position	Variant	Strand	Feature	AA Translation	Frequency (%)	Sample Size	Phenotype	Submitted By	Submitted Position
chr13:94751618	C/T	minus	5' UTR		96.35/3.65	548		PMT	92



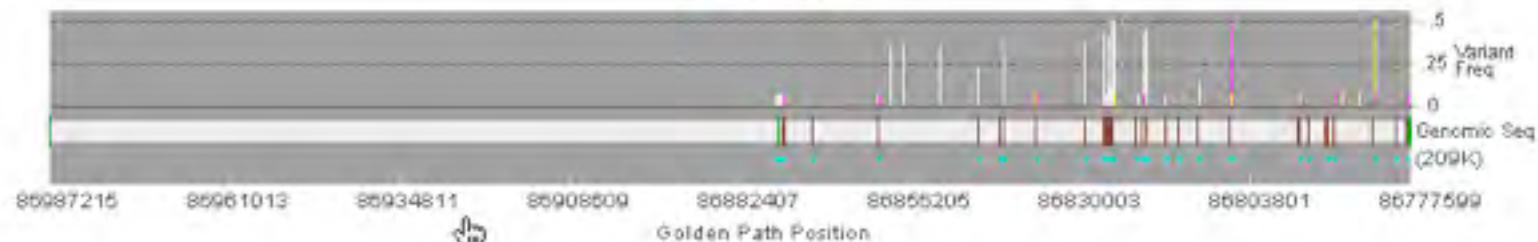
Variant Positions on ABCB1



Web Services
[Comparative Genomics](#)

Gene | ☐ Intron ☒ Exon ☒ UTR ☒ Promoter ☒ Flanking
Variant | ☐ Non-coding/Unknown ☒ Non-synonymous ☒ Synonymous

☒ Assayed Region ?



Click To: ☐ Move ☒ Magnify

Current Position: 86882407

Move:



50



Magnify:



Legend

- deletion

* feature derived from NCBI RefSeq[?]

Number of variant positions: 60



Golden Path Position	Variant	Strand	Feature	AA Translation	Frequency (%)	Sample Size	Phenotype	Submitted By	Submitted Position
chr7:86875319	G/A	minus	Intron*		99.37/0.63	474		PMT	39
chr7:86875268	C/T	minus	Intron*		99.8/0.2	490		PMT	90
chr7:86875191	T/C	minus	Intron*		99.8/0.2	490		PMT	167
chr7:86875105	A/T	minus	Intron*		99.59/0.41	490		PMT	253
chr7:86875086	A/G	minus	Intron*		99.8/0.2	490		PMT	272
chr7:86874956	G/A	minus	5' UTR*		99.79/0.21	480		PMT	98



Variants linked to phenotypes

chr7:86805269	G/T/A	minus	Exon*	Ala/Ser/Thr	60.37/37.96/1.67	598		PAAR-SJCRH	256
chr7:86805269	G/T	minus	Exon*	Ala/Ser	55.71/44.29	438		PAAR-SJCRH	280
chr7:86794712	G/A	minus	Intron		99.8/0.2	494		PMT	205
chr7:86789329	C/G	minus	Exon	Pro/Ala	99.8/0.2	494		PMT	150
chr7:86788154	C/T	plus	Intron*		93.72/6.28	876		PHAT	101
chr7:86785824	G/C	plus	Intron*		93.76/6.24	898		PHAT	101
chr7:86783409	T/C	minus	Exon	Trp/Arg	99.8/0.2	490		PMT	98
chr7:86783310	T/A	minus	Exon	Ser/Thr	95.31/4.69	490		PMT	197
chr7:86783296	T/C	minus	Exon*	Ile/Ile	56.9/43.1	58		COBRA	274
chr7:86783296	C/T	minus	Exon	Ile/Ile	60.15/39.85	522		PMT	211
chr7:86783296	C/T	minus	Exon*	Ile/Ile	53.34/46.66	598		PAAR-SJCRH	202
chr7:86783296	C/T	minus	Exon*	Ile/Ile	51.6/48.4	188		PAAR-SJCRH	833



Column Descriptions

Subject ID	PharmGKB Accession IDs for subjects in the PharmGKB.
Gender	Male, female, or unknown.
Race/Ethnicity	Self-reported information.
Seizures	Seizures, as defined by the National Cancer Institute Common Toxicity Criteria version 2.0 . No = grade 0; yes = grades 1-4.
CNS Thrombosis	Central nervous system (CNS) thrombosis, as defined by the National Cancer Institute Common Toxicity Criteria version 2.0 . No = grade 0; yes = grades 1-4.
Plasma homocysteine at diagnosis (uM)	Plasma homocysteine concentrations (uM) at the time of diagnosis of pediatric acute lymphoblastic leukemia (ALL), measured by HPLC.
Plasma homocysteine, pre-consolidation 1 (uM)	Plasma homocysteine concentration (uM) immediately before course 1 of high-dose methotrexate consolidation therapy, measured by HPLC.
Plasma homocysteine, pre-consolidation 2 (uM)	Plasma homocysteine concentration (uM) immediately before course 2 of high-dose methotrexate consolidation therapy, measured by HPLC.
Plasma homocysteine, consolidation 2, 23hr (uM)	Plasma homocysteine concentration (uM) at 23 hours after the start of course 2 of high-dose methotrexate consolidation therapy, measured by HPLC.
Plasma homocysteine, consolidation 2, 44 hr (uM)	Plasma homocysteine concentration (uM) at 44 hours after the start of course 2 of high-dose methotrexate consolidation therapy, measured by HPLC.

Subject ID	Gender	Race/Ethnicity	Seizures	CNS Thrombosis	Plasma homocysteine at diagnosis (uM)	homocysteine, pre-consolidation 1 (uM)	homocysteine, pre-consolidation 2 (uM)	homocysteine, consolidation 2, 23hr (uM)	homocysteine, consolidation 2, 44 hr (uM)	Plasma homocysteine, pre-week 31 (uM)	CSF homocysteine at diagnosis (uM)
PA126747064	Not specified	Not specified	no	no	4.13	53.84	3.98	9.72	8.54	7.26	0.055
PA126747065	Not specified	Not specified	no	yes	9.34	4.73	5.65	9.22	10.89	6.9	0.043
PA126747066	Not specified	Not specified	no	no	3.24	4.59	4.42	12.28	9.94		0.055
PA126747067	Not specified	Not specified	yes	no	20.06	14.04	5.39				0.044
PA126747068	Not specified	Not specified	no	no	8.44	4.71	7.16	6.95	5.61	9.52	0.106
PA126747069	Not specified	Not specified	no	no		4.2	8.0	12.31	9.94	8.14	
PA126747070	Not specified	Not specified	yes	no	8.54	9.12	3.95				0.095
PA126747071	Not specified	Not specified	no	no	2.16	4.54	5.77	9.91	10.81	6.51	0.035
PA126747072	Not specified	Not specified	no	no	38.92						0.109
PA126747073	Not specified	Not specified	no	no	6.27	9.1	6.05	4.39	4.63	5.13	0.071
PA126747074	Not specified	Not specified	no	no	11.97	3.87	10.11	14.4	5.32	4.3	0.082
PA126747075	Not specified	Not specified	yes	no							0.063
PA126747076	Not specified	Not specified	no	no	6.98	3.91	2.58	2.73	3.76	3.7	0.055
PA126747077	Not specified	Not specified	no	no	9.47	5.1	2.9	4.74	4.46	4.5	0.043



Ethnicity: Not Specified

This subject is a member of the following sample sets:

- [SERM Sample Set PA130054054](#)
- [SERM Sample Set PA128768829](#)
- [SERM Sample Set PA128770662](#)
- [SERM Sample Set PA130150459](#)
- [SERM Sample Set PA130150886](#)
- [SERM Sample Set PA129968859](#)

Genotype Data

Gene	Number of Variant Positions	Data	
ABCB1	1	 View Genotype	 Download Data
CYP2C19	2	 View Genotype	 Download Data
CYP3A5	1	 View Genotype	 Download Data
ESR1	2	 View Genotype	 Download Data
NOS3	3	 View Genotype	 Download Data
		 View All	 Download All



irinotecan

Generic Names: 7-ethyl-10-[4-(1-piperidino)-1-piperidino]**Trade Names:** Ca**PharmGKB
Accession ID:**
PA450085[Add Data](#) **References**

None

on Searches[Search PubMed](#)[Medline Plus](#)[Search PubChem](#)

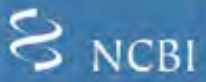
PharmGKB Primary Data

Phenotype Datasets

 [Irinotecan Clinical](#)*Submitted by: Mark J.
TOP1, UGT1A1, UGT1A* [Irinotecan and CYP](#)*Submitted by: Mark J.* [Pharmacokinetics](#)*Submitted by: Howard
CYP3A5, UGT1A1, XRC* [Protein expression](#)*Submitted by: Howard
CYP3A4, CYP3A5, TOP*

Pathways

- [Irinotecan Pathway](#)
- [Irinotecan Pathway \(Cancer\)](#)

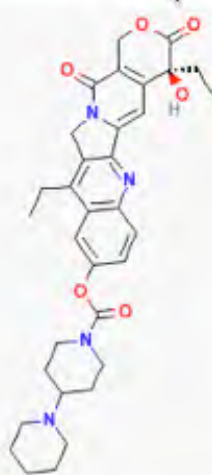
 **PubChem** 
National Library of Medicine

HOME SEARCH CITE MAP PubMed Entrez Structure GenBank PubChem

Search

Substance Summary:

Compound Displayed



Source: [ChemIDplus \(097682445\)](#)

 **SID:** 217212 [?](#)

 **CID:** 60838 [?](#)

 **NLM Toxicology:** [Link](#) [?](#)

 **Related Substances:** [?](#)
Same: [3 Links](#)
Same, Connectivity: [4 Links](#)

 **Similar Substances:** [23 Links](#) [?](#)

 **Structure Search** [?](#)



Literature Annotations

Related Genes	
   ABCB1	
   ABCC1	
   ABCG2	
 BCHE	
   CES1	
   CES2	
   CYP3A4	
   CYP3A5	
   MTHFR	
   SLC19A1	
 SLCO1B1	
   TYMS	
   UGT1A1	

Legend

	Related Diseases	Relationship ?	Details
 Colonic Neoplasms	CO	PK	
 Colorectal Neoplasms		FA	
 Gilbert syndrome		PK	
  Neoplasms	CO	PD PK	GN 
 cancer or viral infections		PK	GN 

Details ✦

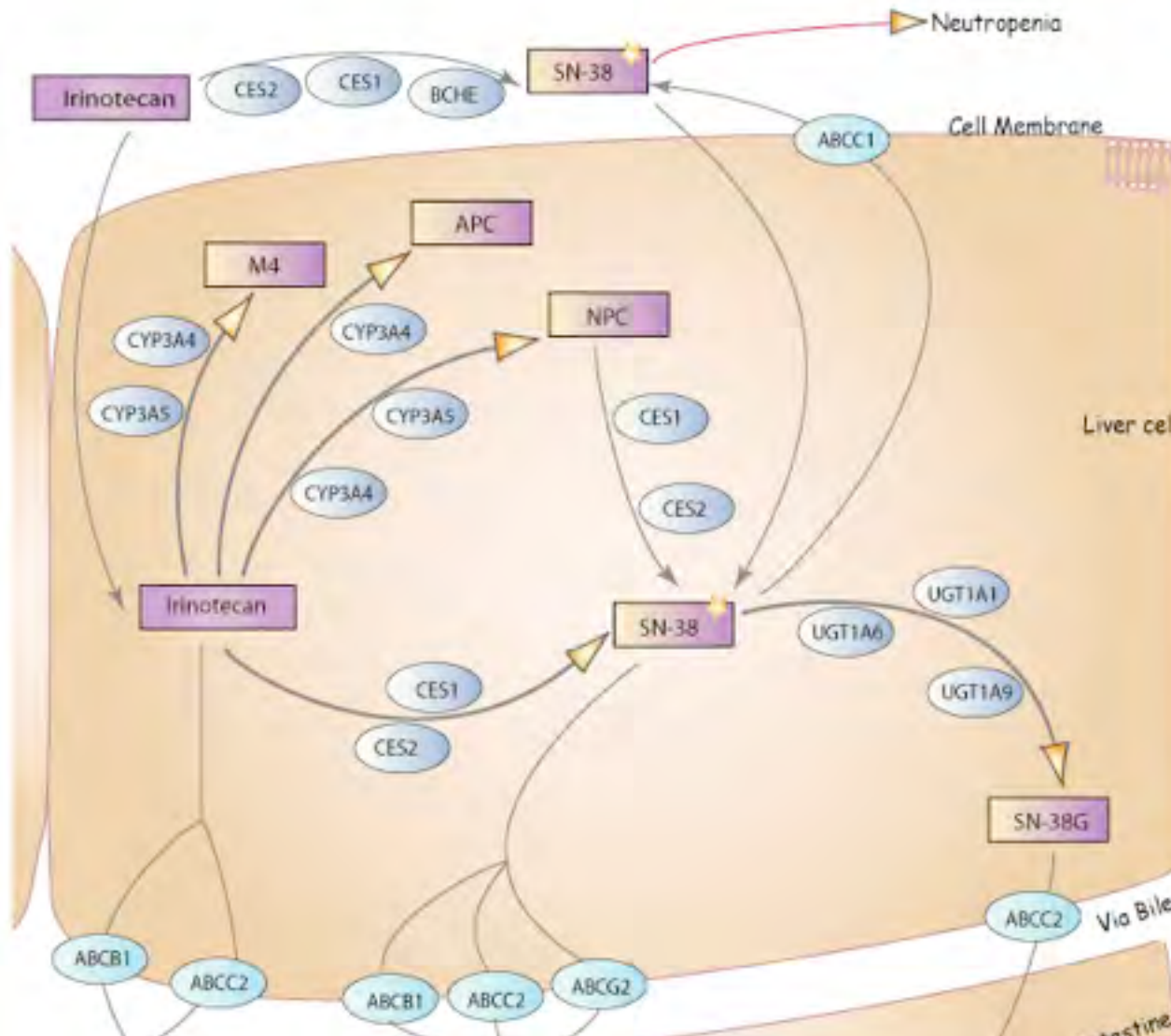
- [Molecular weight](#)
- [Indications](#)
- [Contraindications](#)
- [Mechanisms Of Action](#)
- [Distribution Data](#)
- [Protein Binding Data](#)
- [Biotransformation Data](#)
- [Half Life](#)
- [Peak Serum Concentration](#)
- [Elimination Data](#)
- [Adverse Effects](#)
- [Interactions](#)



Irinotecan Pathway

Liver cell:

Model human liver cell showing blood, bile and intestinal compartments, indicating tissue specific involvement of genes in the irinotecan pathway.



[All Pathways](#)

liver ▼

Go

RELATED GENES

- [NR112](#)
- [ICF1](#)

RELATED DRUGS

- Topotecan

RELATED DISEASES

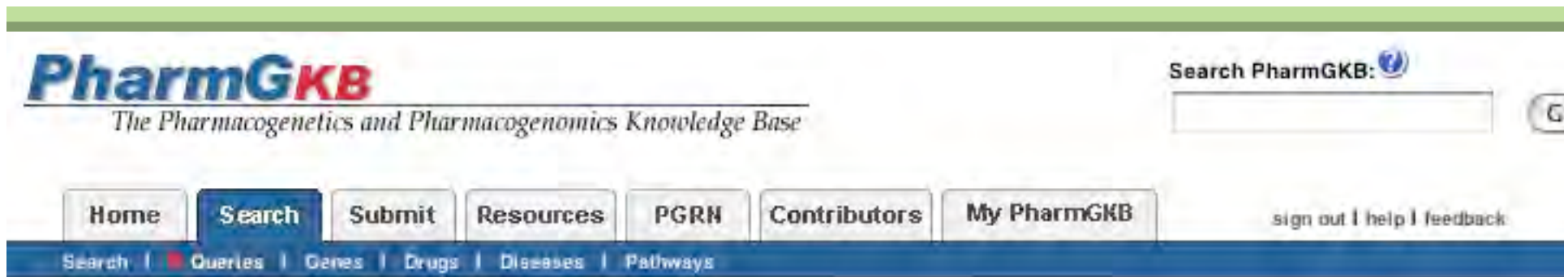
- [Neoplasms](#)
- [Colonic Neoplasms](#)
- [Colorectal Neoplasms](#)
- [Gilbert Syndrome](#)
- [Carcinoma, Small Cell](#)

[DOWNLOADS](#)

- [Illustrator file](#) (liver.ai)
- [Supporting Evidence](#) (xls)



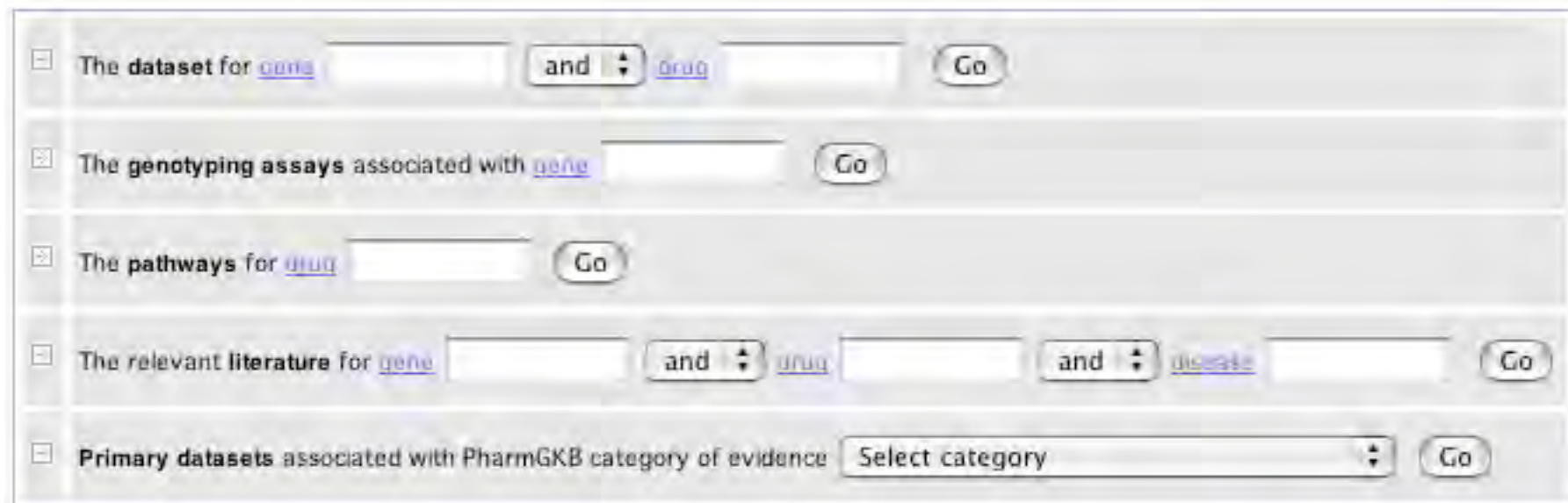
Most common “simple” queries: templates



The PharmGKB website header features the logo "PharmGKB" in blue and red, with the tagline "The Pharmacogenetics and Pharmacogenomics Knowledge Base" below it. A search bar is located in the top right corner. A navigation bar contains buttons for Home, Search, Submit, Resources, PGRN, Contributors, and My PharmGKB. A secondary navigation bar lists Search, Queries, Genes, Drugs, Diseases, and Pathways. A sign out link, help link, and feedback link are also present.

Simple Queries

Please choose the relevant query, select a gene and/or drug and /or disease and press the "Go" button.



The Simple Queries section contains five templates, each with a checkbox, a description, input fields, and a "Go" button. The templates are: 1. "The dataset for gene and drug" with input fields for gene and drug. 2. "The genotyping assays associated with gene" with an input field for gene. 3. "The pathways for drug" with an input field for drug. 4. "The relevant literature for gene and drug and disease" with input fields for gene, drug, and disease. 5. "Primary datasets associated with PharmGKB category of evidence" with a dropdown menu for "Select category".



PA126743803
PA126743804
PA126743805
PA126743806
PA126743807
PA126743808
PA126743809
PA126743810
PA126743811
PA126743812
PA126743813
PA126743814
PA126743815
PA126743816
PA126743817
PA126743818
PA126743819
PA126743820
PA126743821
PA126743822
PA126743823
PA126743824
PA126743825
PA126743826
PA126743827
PA126743828
PA126743829
PA126743830
PA126743831
PA126743832
PA126743833
PA126743834
PA126743835
PA126743836
PA126743837
PA126743838
PA126743839
PA126743840
PA126743841
PA126743842
PA126743843
PA126743844
PA126743845
PA126743846
PA126743847
PA126743848
PA126743849

Genotypes for samples (y-axis) for variants in ACE by Golden Path position (x-axis)

Legend:

-  Homozygous - common
 Homozygous - rare
 Heterozygous
 Non-biallelic
 No data

Phenotype - Genotype Grid Display

Showing all phenotypes with at least one significant correlation to a variant of CYP2B6.

CYP2B6 (Golden Path Position)	PA129840135: CYP2B6 expression and activity			PA646600: Midazolam catalytic activity in liver microsomes			
	CYP2B6 activity (pmol/min/mg)	CYP2B6 Protein Expression (pmol/mg)	Relative CYP2B6 mRNA Expression	Midazolam Activity 4-OH (nmol/mg protein/hr) Measurement 1	Midazolam Activity 1-OH (nmol/mg protein/hr) Measurement 1	Midazolam Activity 4-OH (nmol/mg protein/hr) Measurement 2	Midazolam Activity 1-OH (nmol/mg protein/hr) Measurement 2
chr19:46170940							
chr19:46173323							
chr19:46185999							
chr19:46188841							
chr19:46198500							

Legend:



P < 0.01

P < 0.05

not significant

not tested

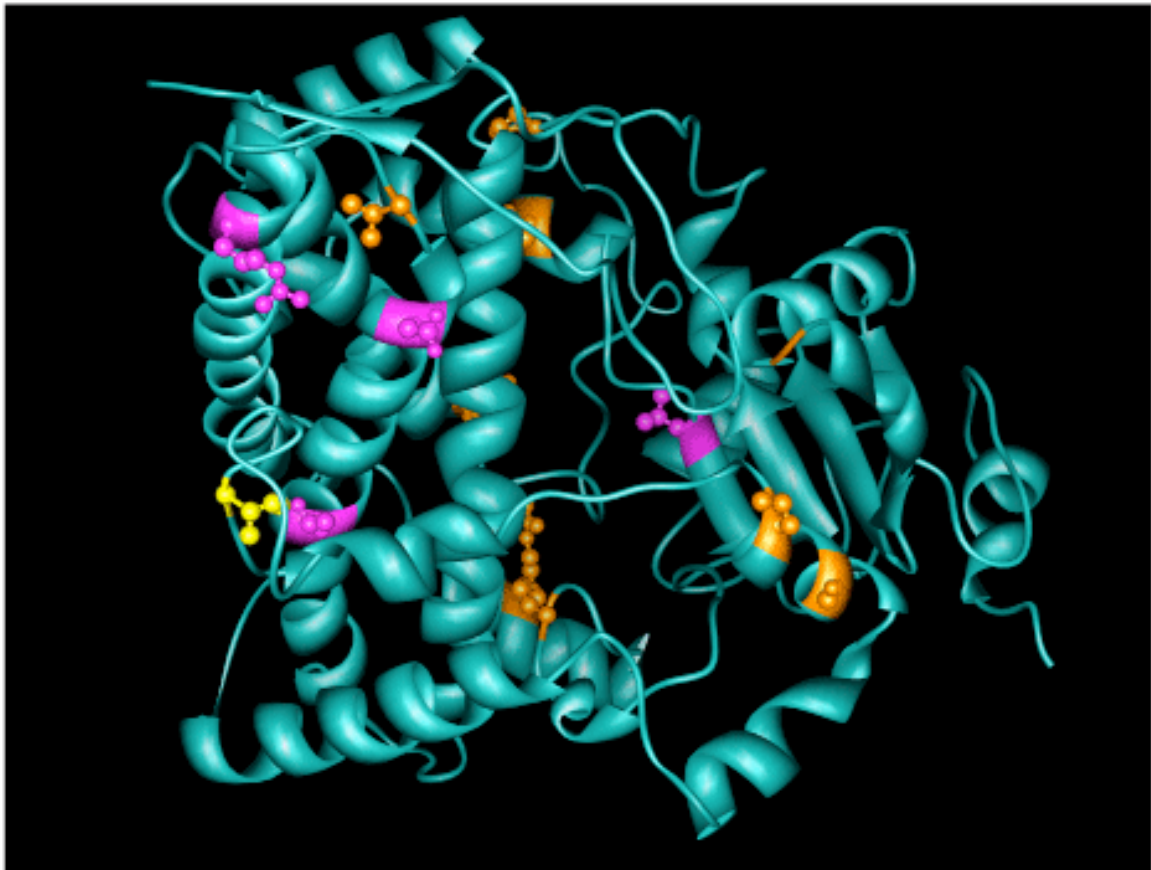
Variant Positions on CYP3A4

- Non-synonymous
- Synonymous
- SNP data reported from external resources



Web Services

[PHASE Analysis](#)
[Comparative Genomics](#)



[PDB: 1WOF](#)

Legend: - deletion * feature derived from NCBI RefSeq

PharmGKB Primary Data

Golden Path	Variant	Strand	Feature	AA	Frequency	Sample	Submitted	Submitted
RefSeq				Translation	100%	Size	PharmGKB	RefSeq

Extended text search

- Boolean combinations AND, OR, NOT: tmpt NOT mercaptopurine
- Wildcards ? and *: tpm*
- Fuzzy match: codine~
- Boosting: CYP3A4 leukemia^4
- Word neighborhood: “TPMT mercaptopurine”~10

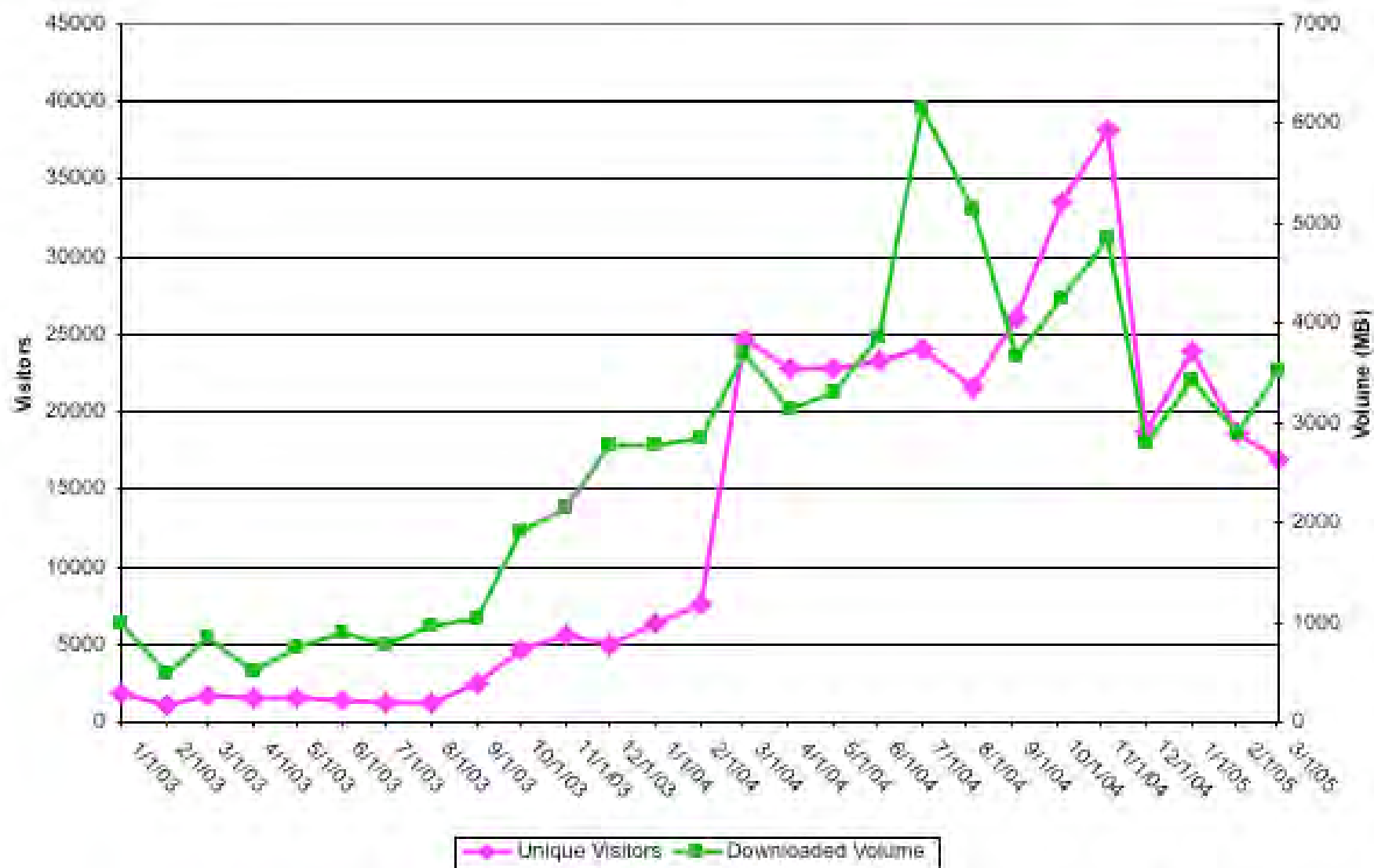


Statistics (thru 3/31/05)

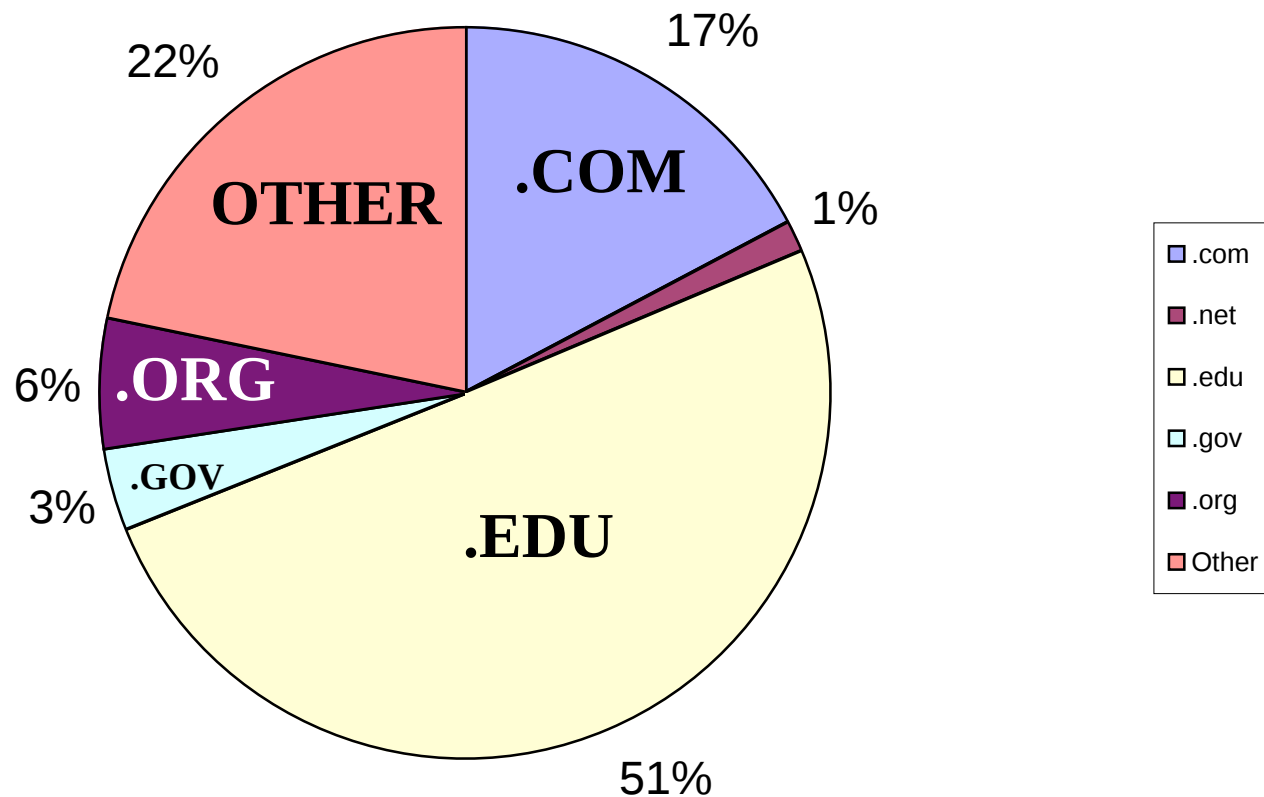
- 1286 registered users (100/month, 50% .edu)
- 1233 literature annotations
- 10578 unique SNP positions (1,008,609 total from 10154 individuals)
- 109649 phenotype measurements
- 5725 subjects with genotype & phenotype data
- 258 genes with data
- 269 drugs with data
- 74 phenotype data sets
- 16 annotated pathways
- 2,166,899 genotype-phenotype pairs



Use of PharmGKB 1/03-3/05



User profiles



Top search terms 10/04-2/05

TPMT (124)

Chemotherapy (82)

cyp2d6 (79)

abcb1 (73)

codeine (57)

irinotecan (52)

ugt1a1 (44)

leukemia (44)

mdr1 (44)

cyp2c19 (38)

warfarin (35)

cyp2c9 (35)

primary (32)

heparin binding protein (32)

drugs (32)

genes (31)

pharmacogenomics (31)

cyp3a5 (29)

ace (29)

methotrexate (28)

insulin (27)

folate pathway (27)

abcg2 (26)

cyp3a4 (26)

pharmacogenetics (26)

variation (24)

cyp2b6 (24)

death (22)

tamoxifen (22)

statin (21)

cisplatin (21)

depression (21)

asthma (21)

Alzheimer (20)

heart failure (20)

nat2 (20)

cyp (20)



Costs of sharing data?

- Need to learn how to do submissions
- Need to take time & resources to format data for submissions
- Worry that competitors will get some advantage with access to actual data



Benefits of sharing data?

- Benefits for the field
- Benefits for the individual
- Benefits for others



The field of pharmacogenomics (I)

- A **central focal point** for access to pharmacogenomic information creates a presence for the field to attract attention widely
- Data sharing reduces unnecessary duplication
- Faster dissemination of data/methods -> faster progress in the field
- Opportunity to aggregate data to test hypotheses with greater power.



The field of pharmacogenetics (II)

- Standards will develop as a result of sharing
 - Standards of evidence
 - Standards for data exchange
 - Standards for “best practices” in lab
- Available data sets stimulate new analytic/informatic methods.
- Creates “bird’s eye” view of the field for someone entering, catalyzes new ideas.



The individual scientist: **IMPACT**

- More visibility for your work
 - Access 20K+ unique visitors/month
 - Much wider group of scientists (entering from different perspectives--pathways, drugs, genes, literature)
- Increased links to your papers
- Entering point for potential collaborators
- ??Increases chance of paper acceptance
 - Can mention submission in editor letter
 - We can arrange early access by reviewers, if appropriate



1. Start su
submissi
2. PhD-level
3. Web-ba
4. Spreads
5. PREVIEW
6. Ensure l
pointing
7. Simultar
accessio

Contributors | My PharmGKB | sign in | **help** | feedback

Team | Register

on in human genes

[PharmGKB Primary Data](#)
[PharmGKB Primary](#)

[Browse](#)

Common questions

News

- [PharmGKB Newsletter July 29, 2005](#)
- New Data: [Variation in VKORC1](#) and [warfarin dosing](#), reported in the June 2 issue of the [NEJM](#)
- New Feature: [Simple Queries](#)
- New Feature: Named Alleles (e.g. [CYP2C9*2](#))
- New Feature: New Drug Page (e.g. [methotrexate](#))
- [Biomedical Beat](#): A monthly digest of research news from NIGMS.
- [Genotype Excel Templates](#) (updated 5/28/05)



Other interests

- Funding agencies: clear movement toward data sharing, in order to have tangible result of funding decisions.
 - NIH in US now routinely adding data sharing language to grant awards
- Journals: clear movement towards data deposition associated with publication
 - PharmGKB is convening panel of editors to advise.



Costs of sharing data?

- Need to learn how to do submissions
 - Curators are motivated to make this as easy as possible.
- Need to take time & resources to format data for submissions
 - Synchronization with submission is best time when data is “fresh” in your mind.
- Competitors will get some advantage with access to actual data
 - No examples yet of damage
 - Benefits of increased impact accrue nonetheless.
 - Curators can work to release elements of data set as work is published.



Five technical informatics issues

1. Overall architecture
2. RDBMS vs. KB
3. XML for data exchange
4. UI considerations
5. Pathway representation issues
6. Terminologies (genes, drugs, diseases)
7. Makeup of PharmGKB team



Thanks.
russ.altman@stanford.edu

<http://www.pharmgkb.org/>

Director: Teri Klein

Curators: Michelle Carrillo, Caroline Thorn, Andrew MacBride, Mark Kiuchi, Daniel Holbert, James Gallerani, Ira Klein

Technical Staff: Mark Woon, Mei Gong, Winston Gor, Feng Liu, Tina Zhou, TC Truong, Tiffany Jung

