

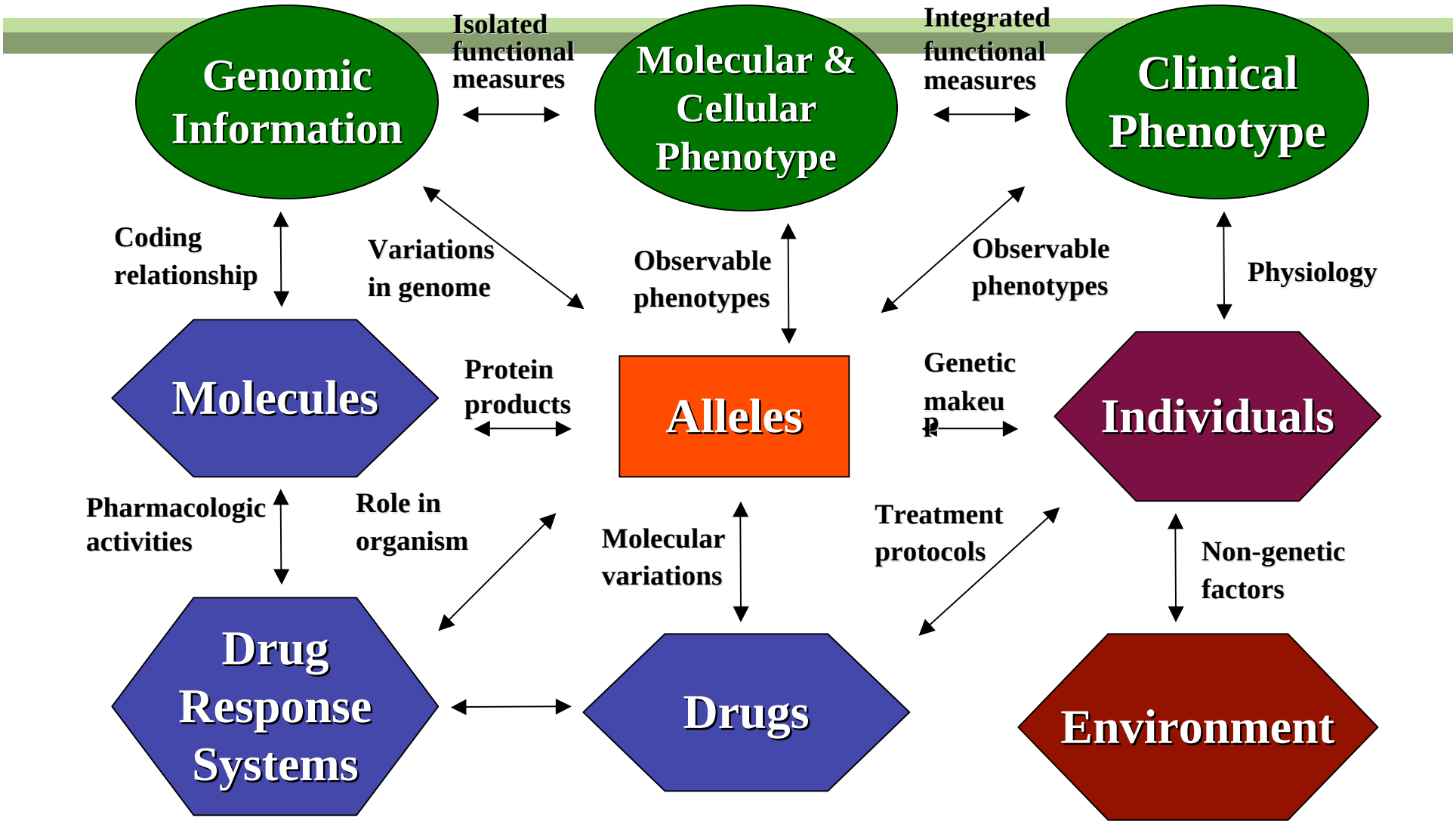
# Building the PharmGKB: a web tool for pharmacogenomics

Russ B. Altman, MD, PhD

Professor of Genetics, Bioengineering & Medicine

Stanford University

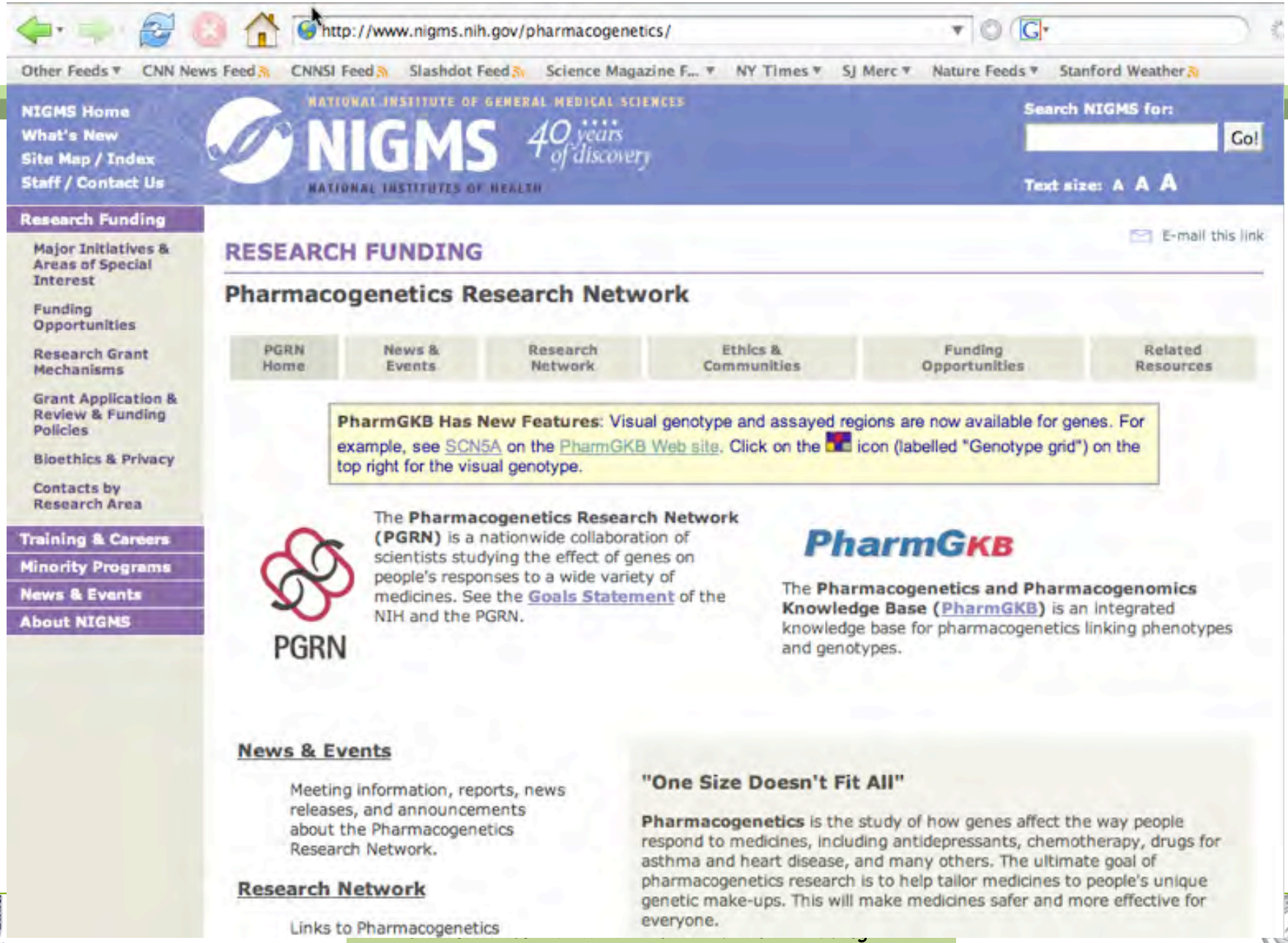
# Complexity of Relationships in Pharmacogenetics



# 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/



# 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

**Arrythmia 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

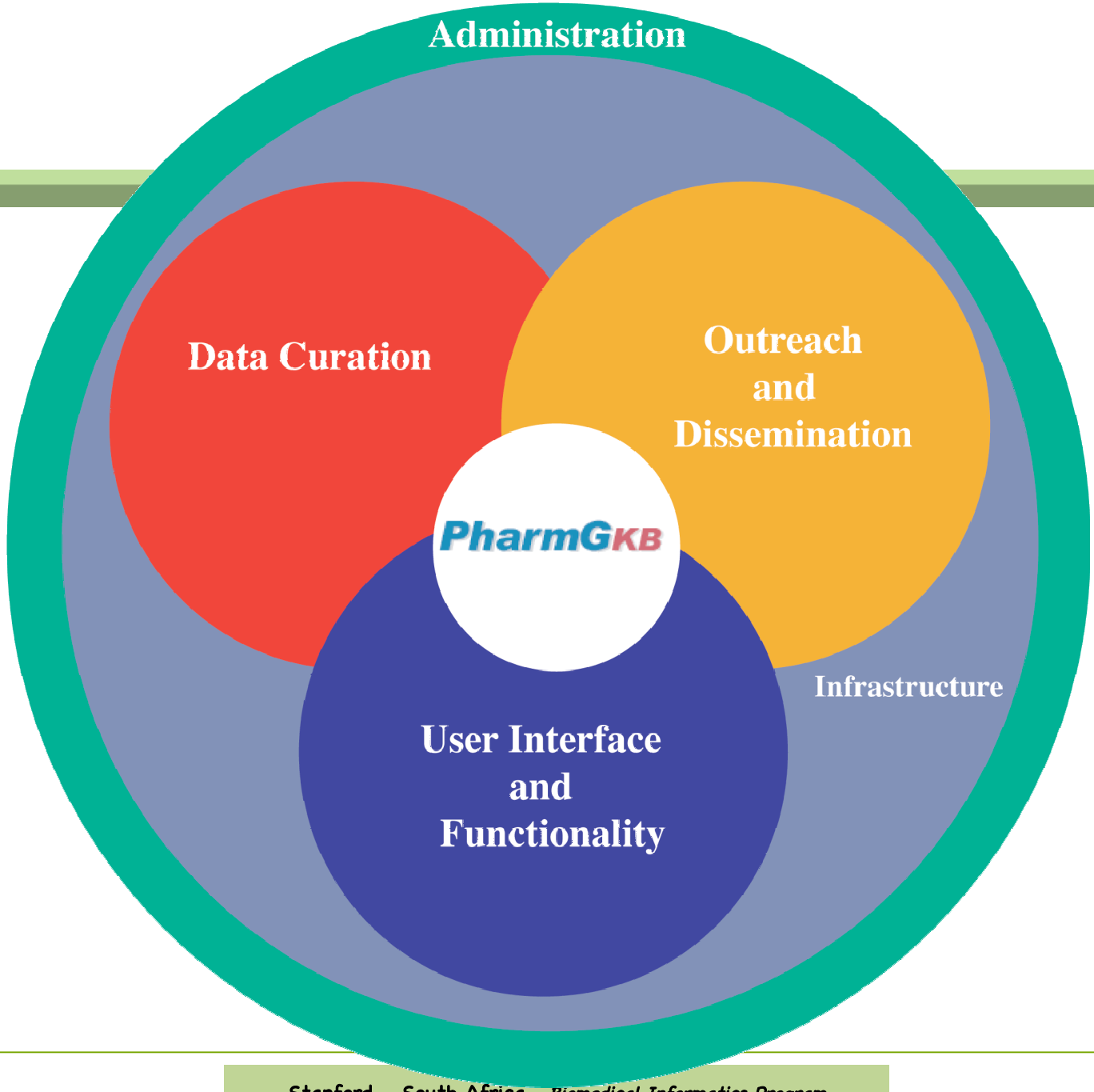
**Phase II Drug Metabolizing Enzymes**
R. Weinshilboum
Mayo Foundation

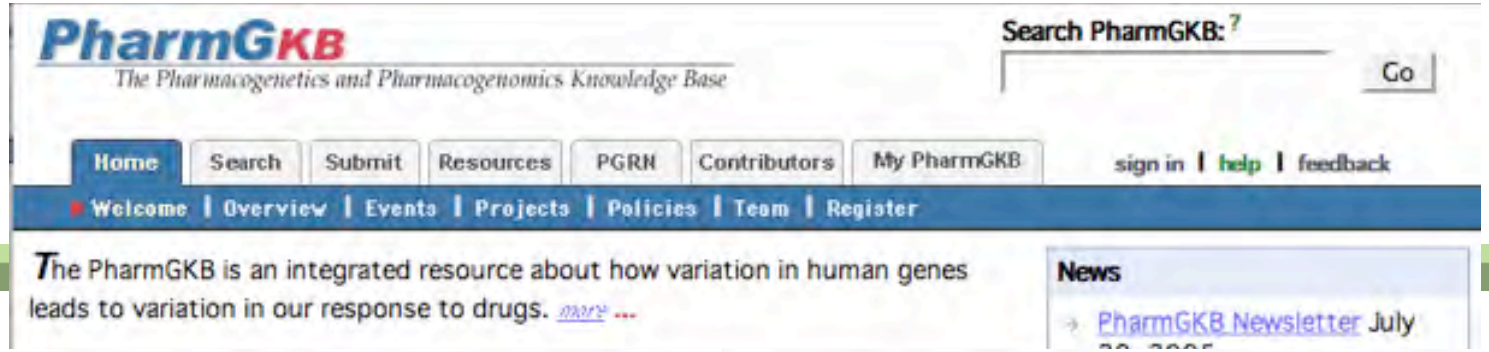
**Nicotine Addiction**
N. Benowitz
University of California, San Francisco

# PharmGKB Goals

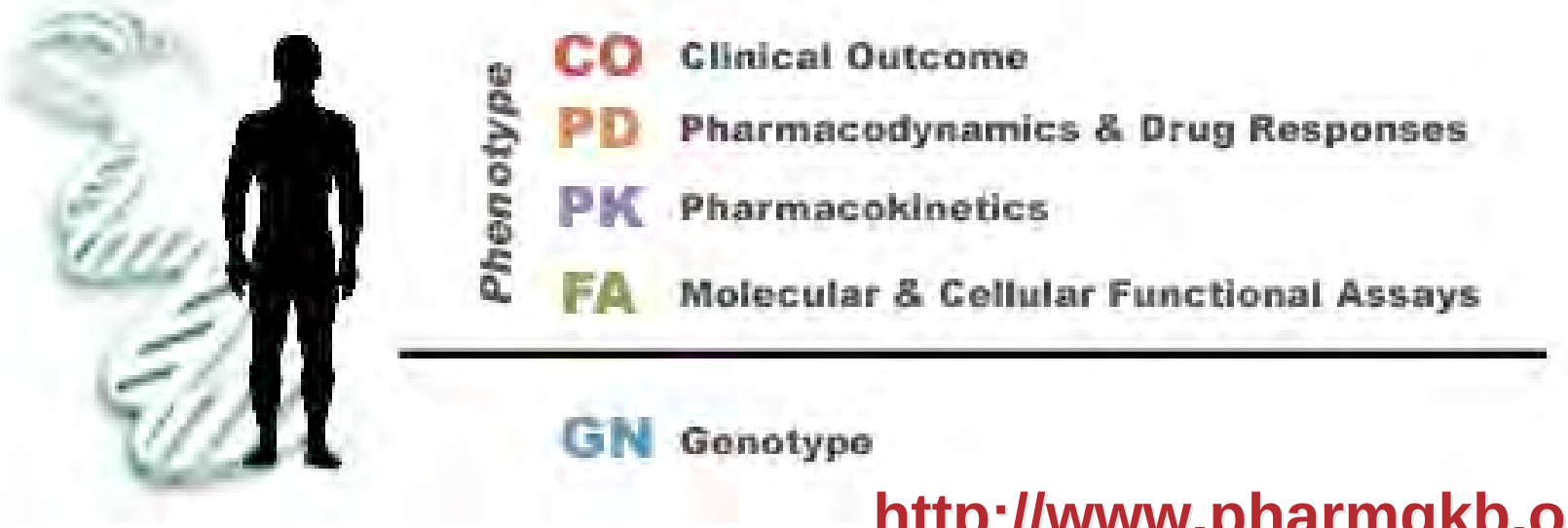
- International data resource of high quality data linking genotypes, phenotypes, pathways, and literature.
- Curate data and develop exchange standards for pharmacogenetics and pharmacogenomics
- Provide analytic functionality for users
- Link with complementary databases such as Medline, PDB (Protein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



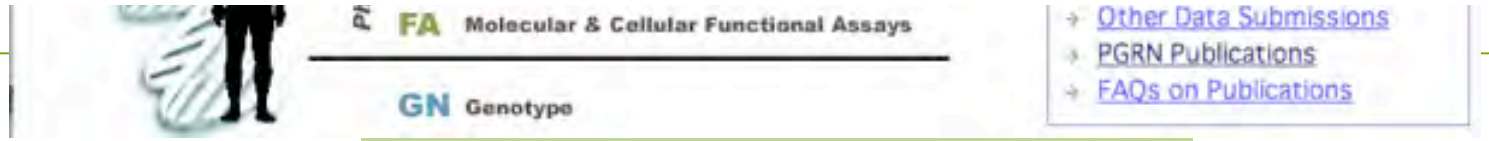

Administration
Data Curation
Outreach
and
Dissemination
PharmGKB
Infrastructure
User Interface
and
Functionality




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**



PharmGKB
The Pharmacogenetics and Pharmacogenomics Knowledge Base

Search PharmGKB: Go

Home | Search | Submit | Resources | PGRN | Contributors | My PharmGKB

sign out | help | feedback

## NOS3
**nitric oxide synthase 3 (endothelial cell)**

| | |
|---|---|
| **Alternate Names:** | ECNOS; endothelial nitric oxidase synthase; endothelial nitric oxide synthase |
| **Alternate Symbols:** | ECNOS; NOS III; eNOS |

Genotype data

PharmGKB Primary Data

**Variant Positions:**

121 positions submitted

Phenotype data

**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 Flockh... CYP2C9, CYP2D6, CYP3A5, ESR1, NOS3, SULT1A1, UGT1A6, tamoxifen, and Br...*

Pathway data

**Pathways:**

- ACE-inhibitor Pathway
- VEGF Pathway

**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:

Literature annotation data

genotype data available

literature annotations available

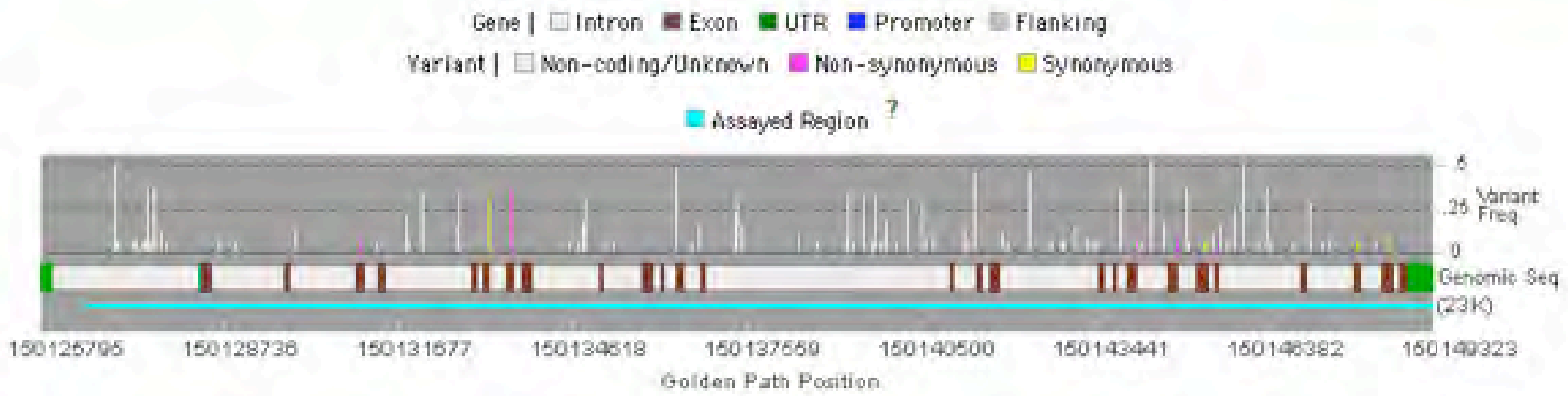
Literature Annotations

| | | 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 | CO PD GN | |
| | | prostaglandins | PD GN | |
| | | rosuvastatin | PD | |
| | | sildenafil | FA | |
| Φ | | tamoxifen | CO PD PK FA | |

# Variant Positions on NOS3

Genotype grid

**Web Services**
Comparative Genomics




**Click To:** Move Magnify

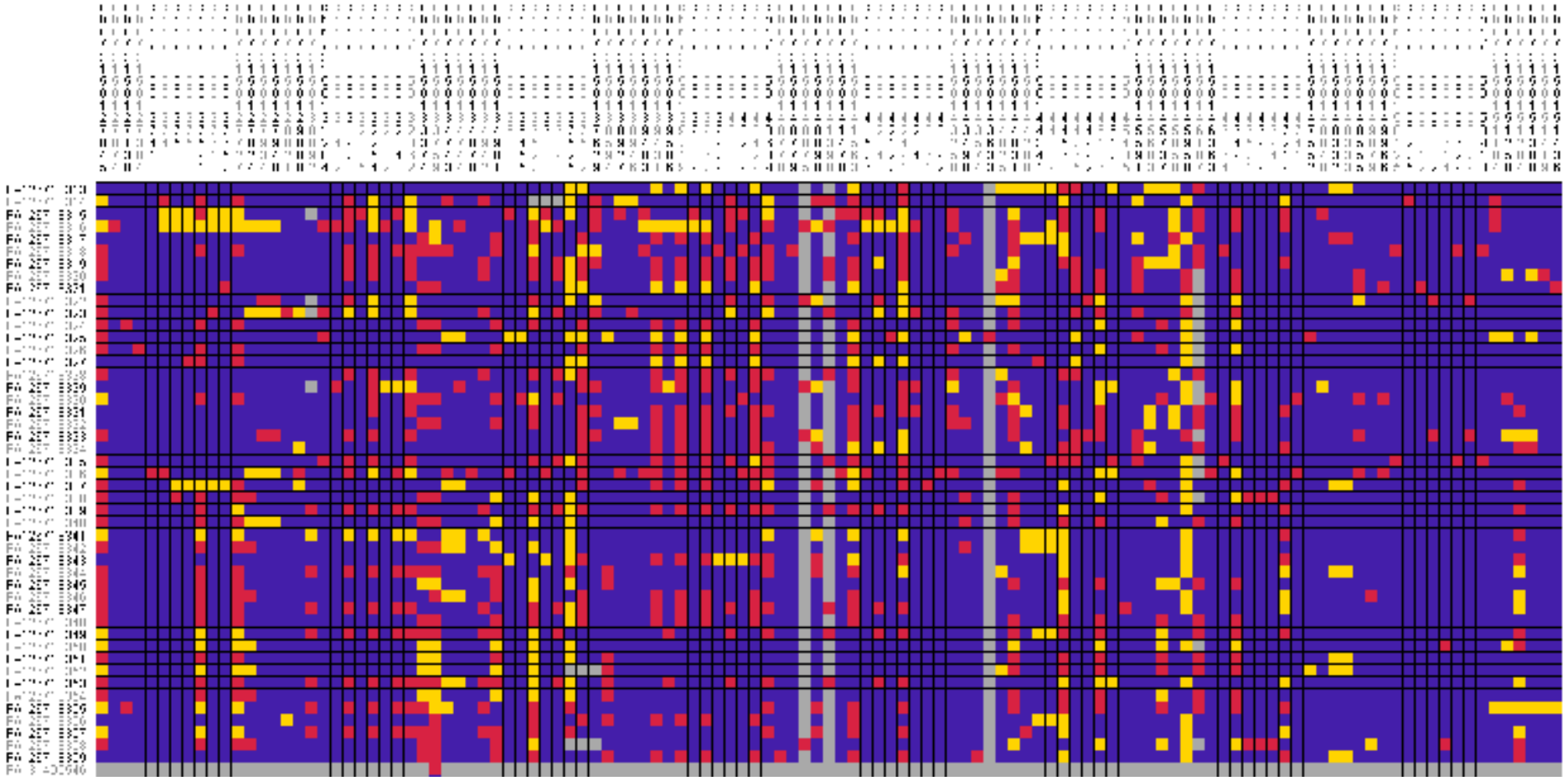
**Current Position:** 150137559 **Move:** < 50 > **Magnify:**

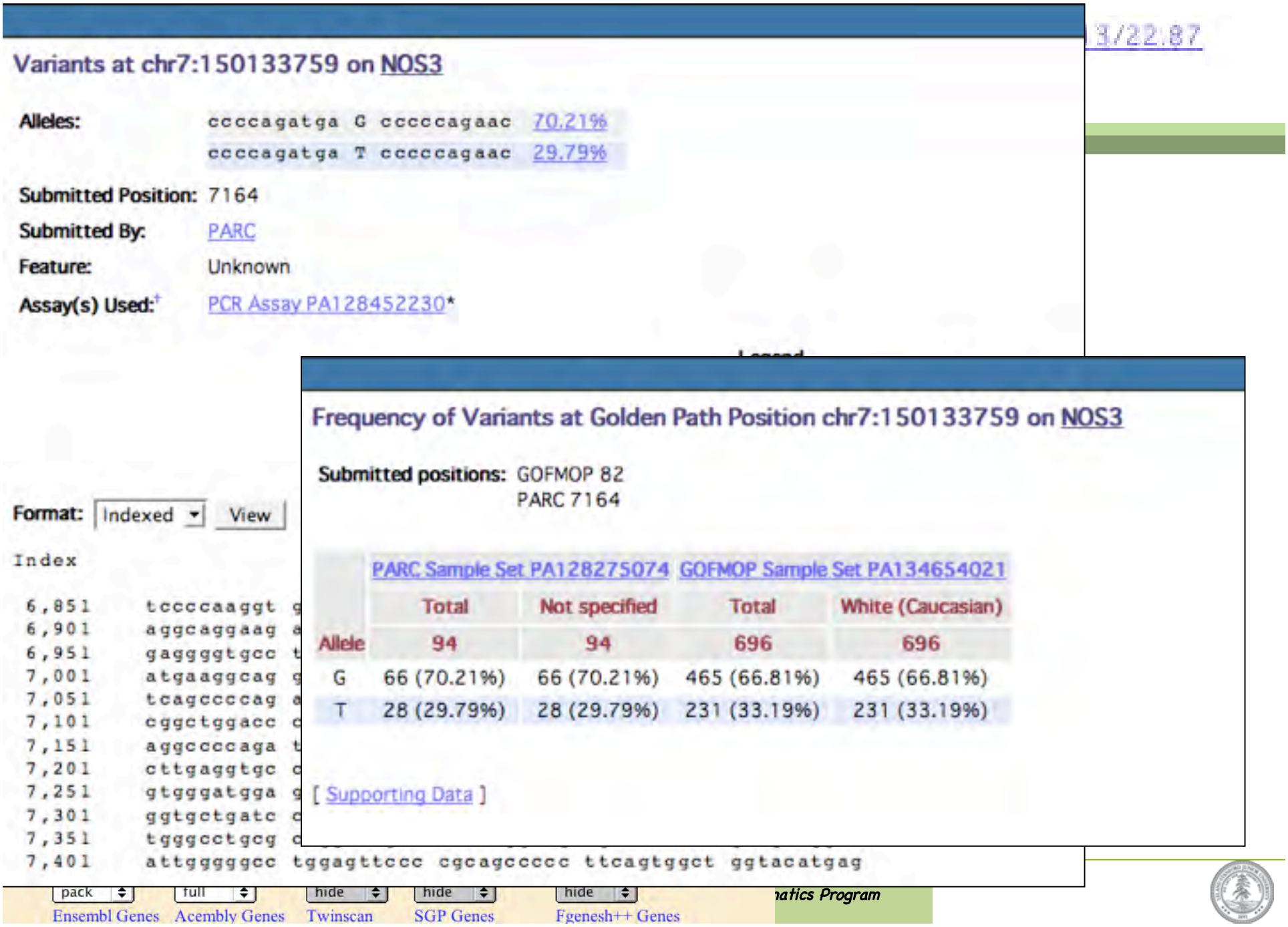
**Legend**
- deletion
* feature derived from NCBI RefSeq†

**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 |

# NOS3 visual genotype

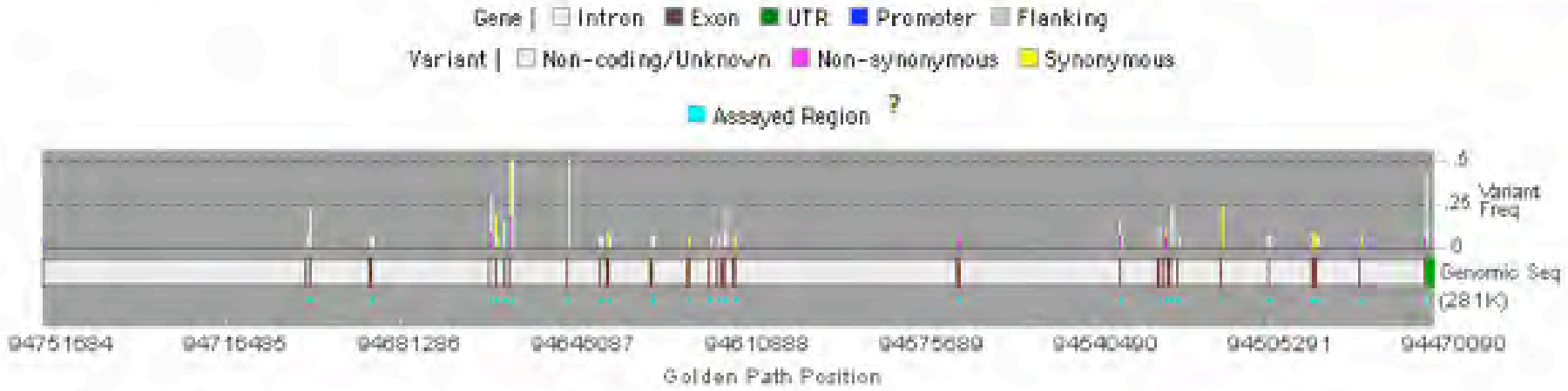



3/22/87
Variants at chr7:150133759 on NOS3
Alleles: ccccagatga G cccccagaac 70.21%
ccccagatga T cccccagaac 29.79%
Submitted Position: 7164
Submitted By: PARC
Feature: Unknown
Assay(s) Used: PCR Assay PA128452230*
Frequency of Variants at Golden Path Position chr7:150133759 on NOS3
Submitted positions: GOFMOP 82
PARC 7164
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 ]
Format: Indexed View
Index
6,851 tccccaaggt g
6,901 aggcaggaag a
6,951 gaggggtgcc t
7,001 atgaaggcag g
7,051 tcagccccag a
7,101 cggctggacc c
7,151 aggccccaga t
7,201 cttgaggtgc c
7,251 gtgggatgga g
7,301 ggtgctgatc c
7,351 tgggcctgcg c
7,401 attgggggcc tggagttccc cgcagccccc ttcagtggct ggtacatgag
pack
full
hide
hide
hide
Ensembl Genes
Acembly Genes
Twinscan
SGP Genes
Fgenesh++ Genes
atics Program

## Variant Positions on ABCC4

Genotype grid

**Web Services**
Comparative Genomics




**Click To:** Move Magnify

**Current Position:** 94610887 **Move:** < 50 > **Magnify:** ?

**Legend**
- deletion
* feature derived from NCBI RefSeq

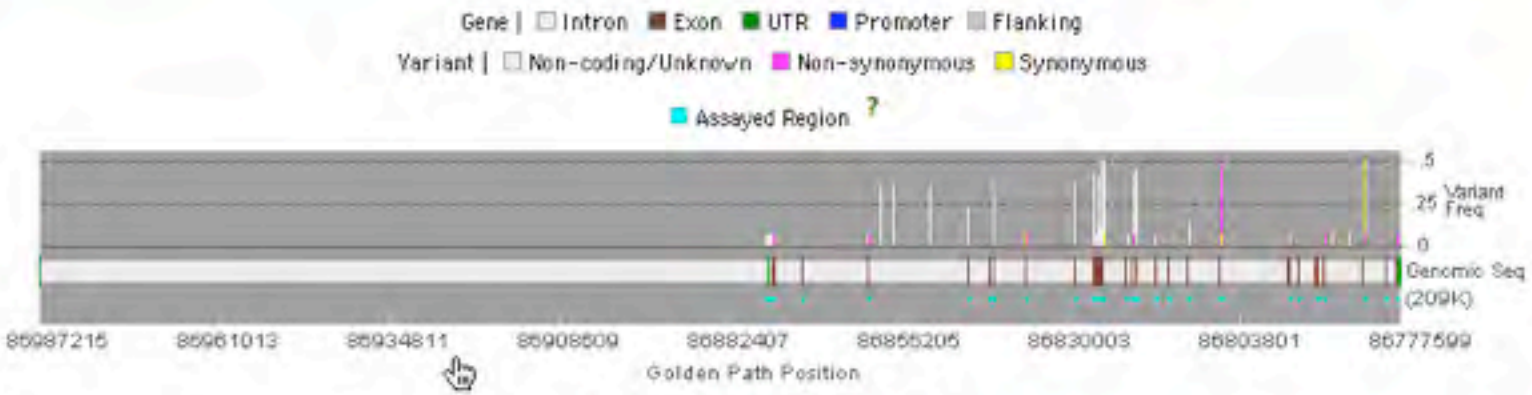
**Number of variant positions:** 99
?

| 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

Genotype grid

**Web Services**
Comparative Genomics




**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 informaiton. |
| **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 homocysteir 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 |

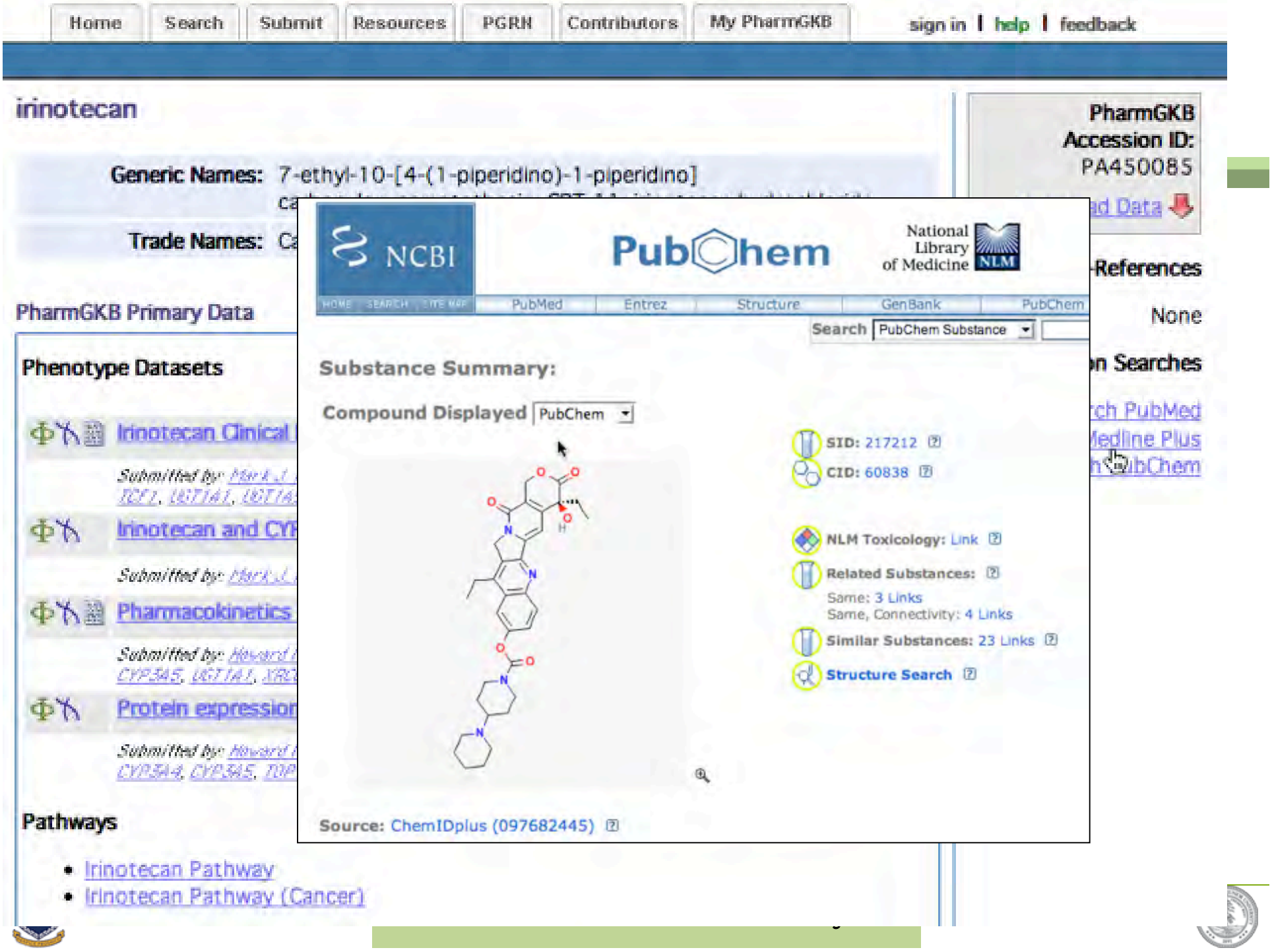
**Ethnicity:** Not Specified

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 |


Home
Search
Submit
Resources
PGRN
Contributors
My PharmGKB
sign in | help | feedback
irinotecan
Generic Names: 7-ethyl-10-[4-(1-piperidino)-1-piperidino]
Trade Names:
PharmGKB Accession ID: PA450085
PharmGKB Primary Data
Phenotype Datasets
Irinotecan Clinical
Irinotecan and CYP
Pharmacokinetics
Protein expression
Pathways
Irinotecan Pathway
Irinotecan Pathway (Cancer)
References
None
Searches
NCBI
PubChem
National Library of Medicine
PubMed
Entrez
Structure
GenBank
PubChem
Search PubChem Substance
Substance Summary:
Compound Displayed PubChem
SID: 217212
CID: 60838
NLM Toxicology: Link
Related Substances:
Same: 3 Links
Same, Connectivity: 4 Links
Similar Substances: 23 Links
Structure Search
Source: ChemIDplus (097682445)

Legend

## Literature Annotations

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

| 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 | |

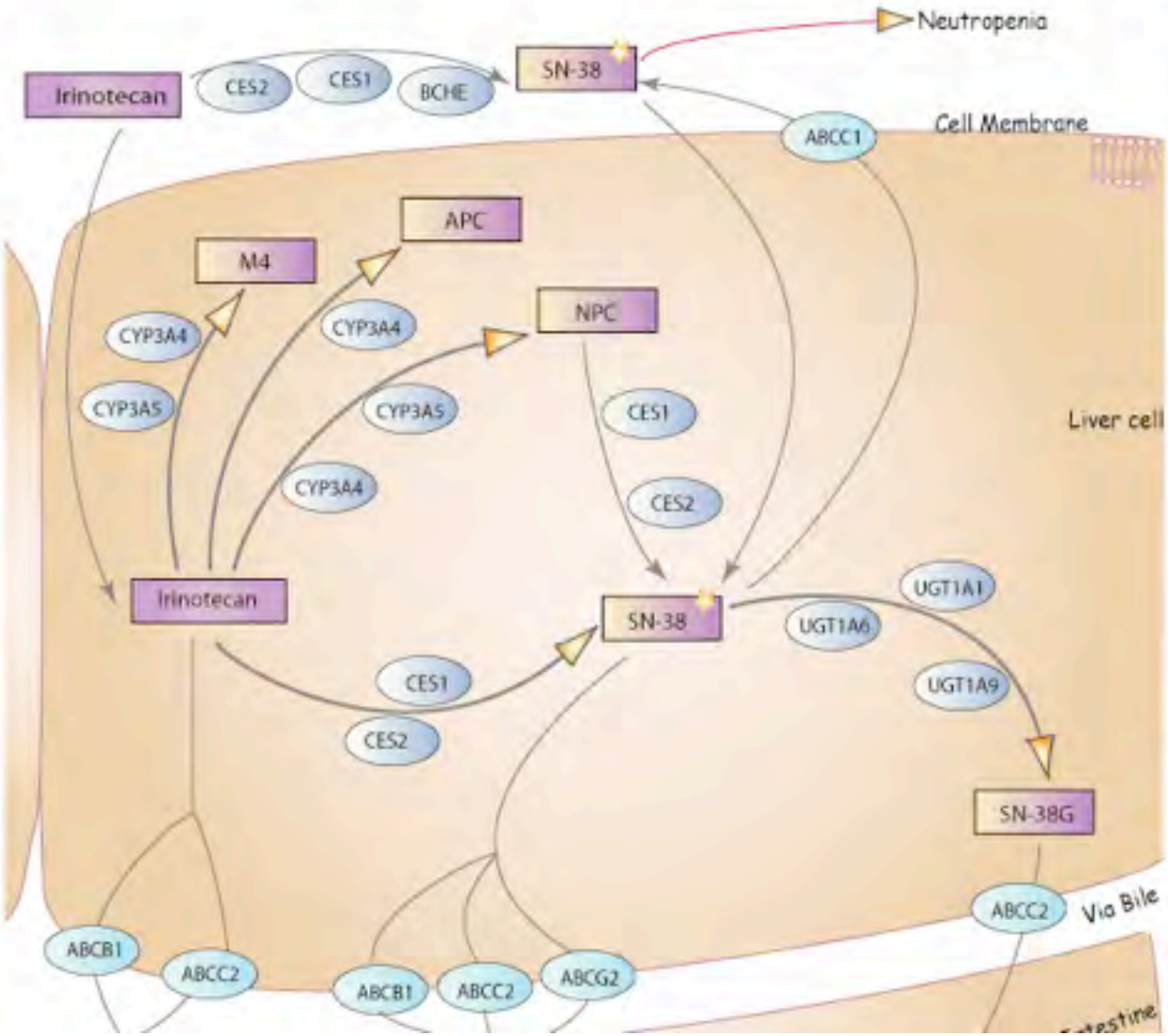
Details

- 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 involvment of genes in the irinotecan pathway.




All Pathways

Legend

liver

Go

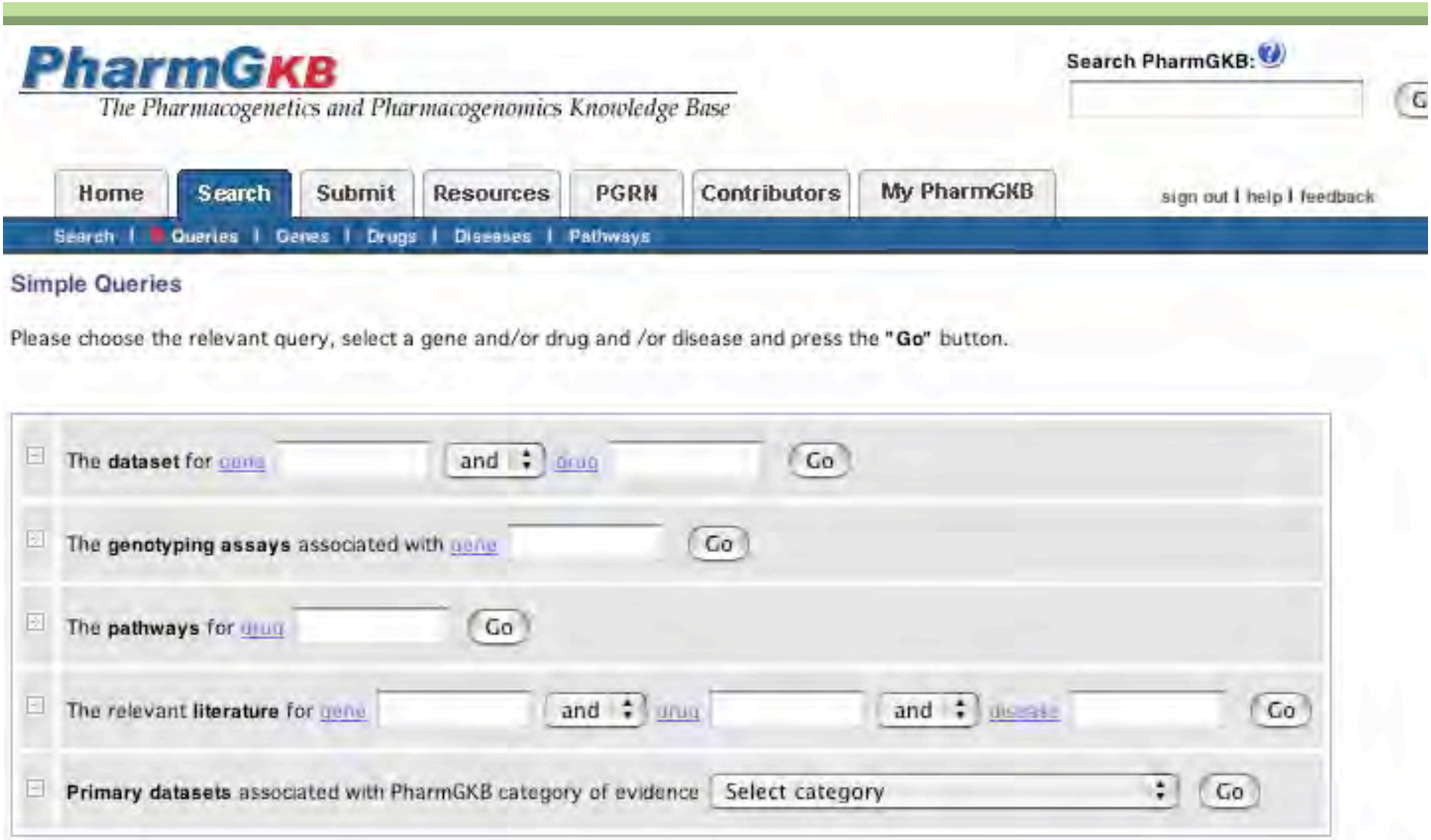
RELATED GENES
- NR1I2
- TCF1

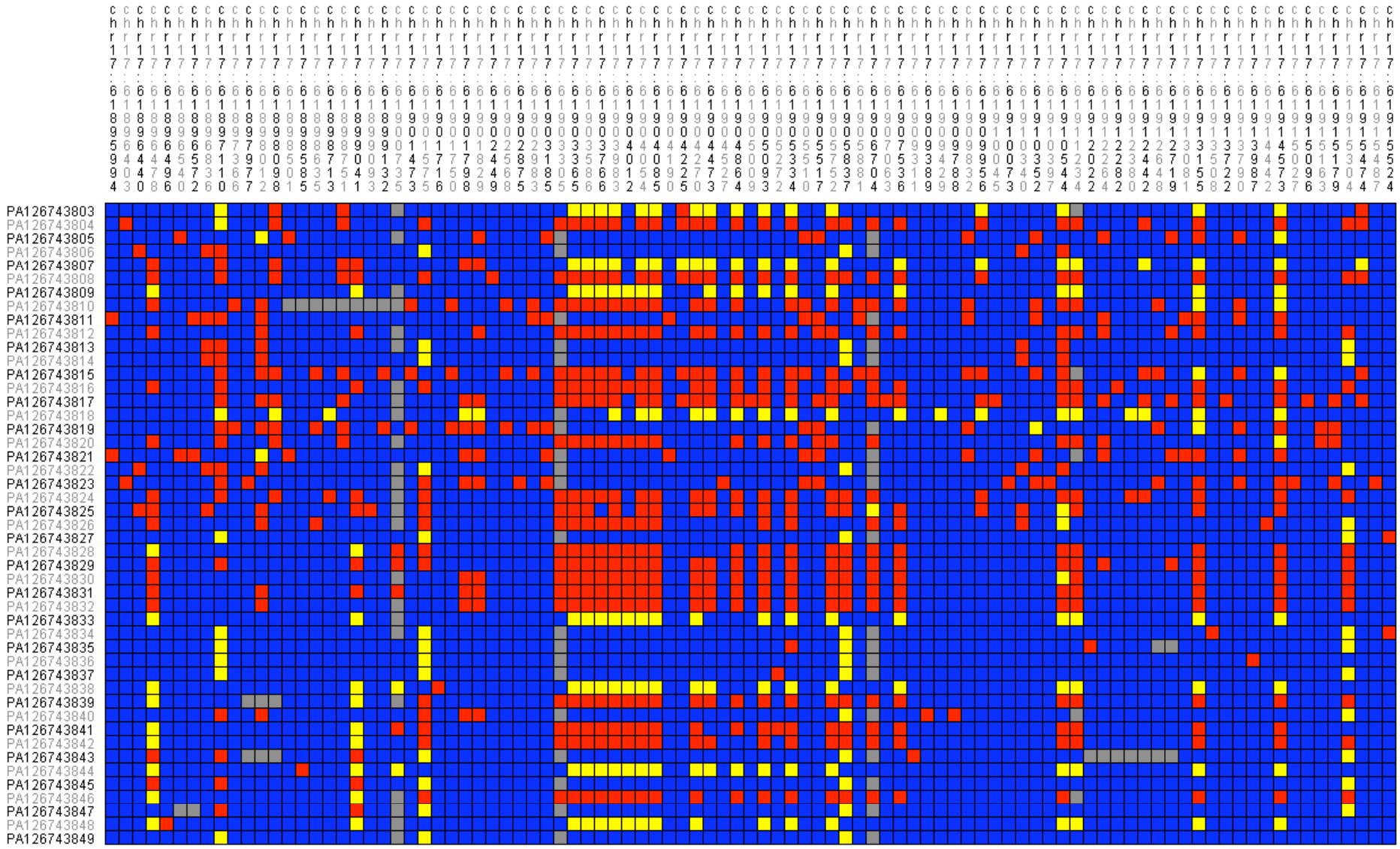
RELATED DRUGS
- Topotecan

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






**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



The PGRN is financially supported by grants from NIGMS, NHLBI, NHGRI, NIEHS, NCI, and NLM within the NIH, HHS. PharmGKB is managed at Stanford University. This work is supported by the NIH/NIGMS Pharmacogenetics Research Network and Database (U01GM61374). 

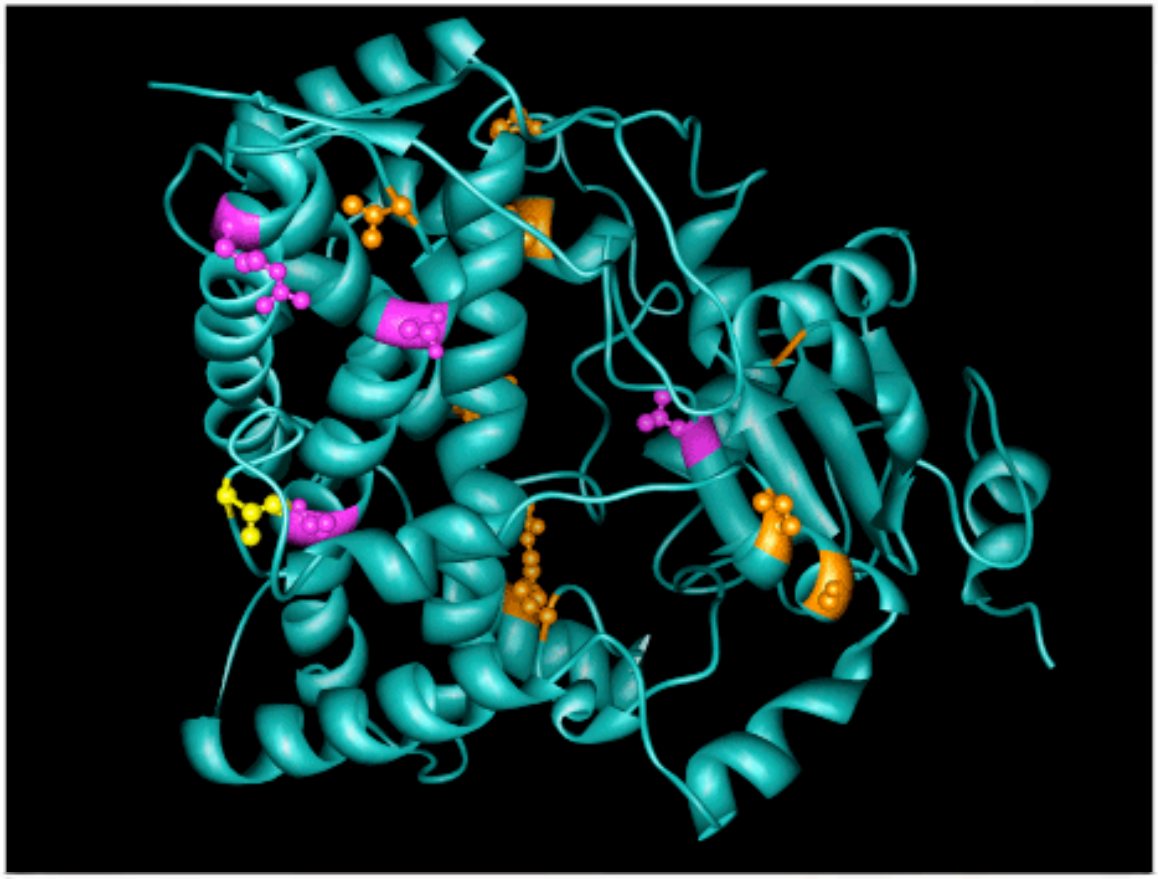
PGRN

## 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 Position | Variant | Strand | Feature | AA Translation | Frequency (%) | Sample Size | Phenotype | Submitted By | Submitted Position |
|---|---|---|---|---|---|---|---|---|---|

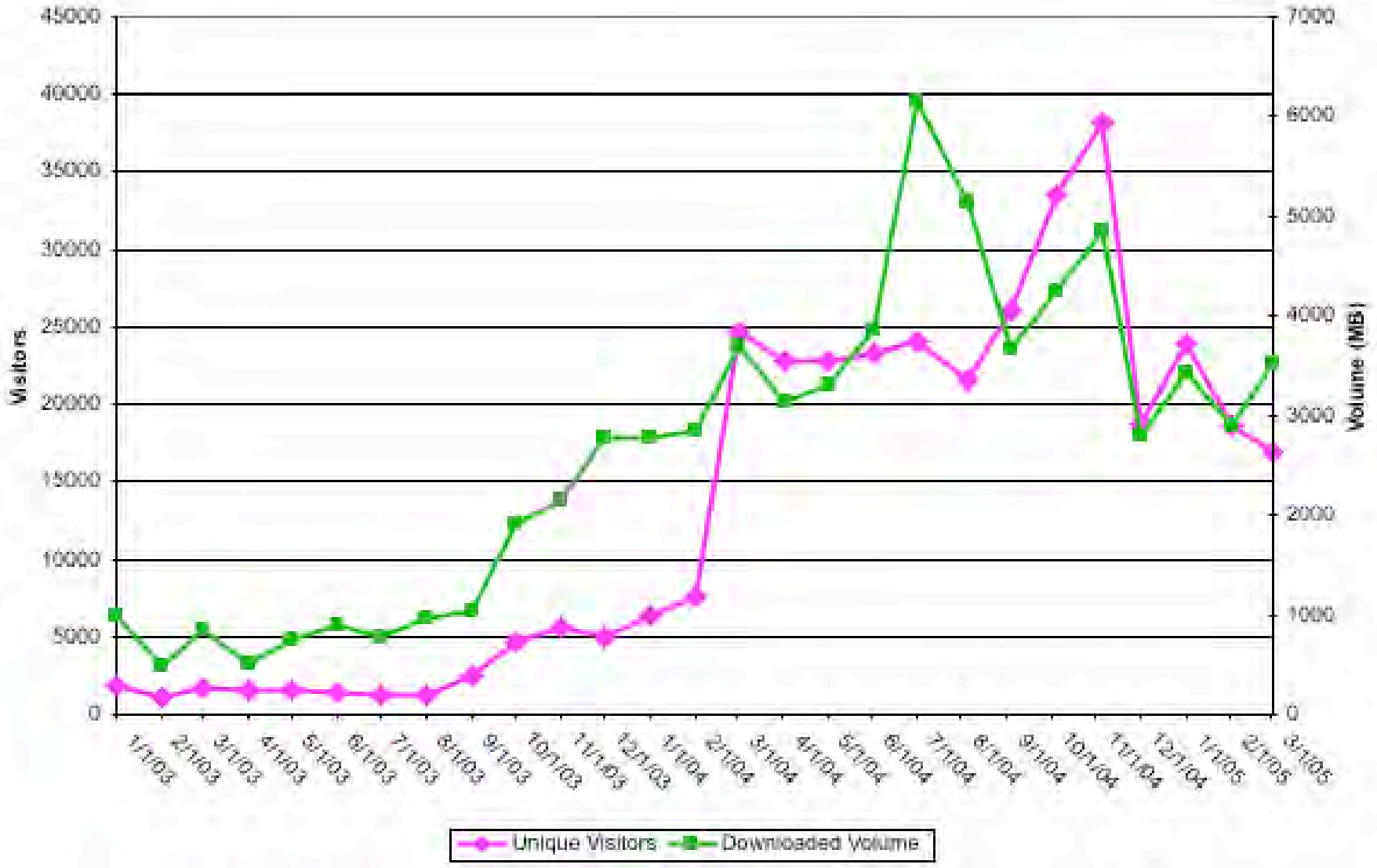
# Extended text search

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

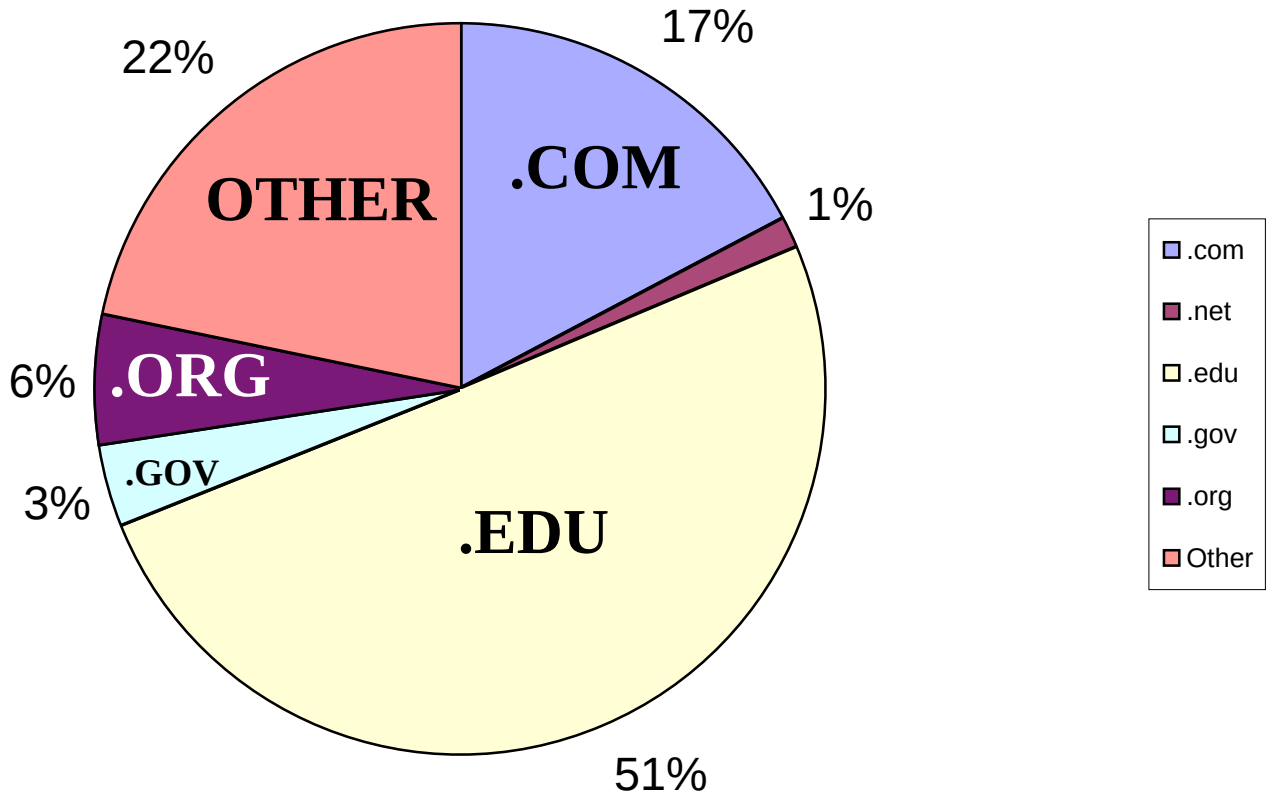
# Statistics (thru 3/31/05)

- 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-leve
3. Web-ba
4. Spreads
5. PREVIEV
6. Ensure l
   pointing
7. Simultar
   accessio

# 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