DIANE E. OLIVER, DANIEL L. RUBIN, JOSHUA M. STUART, MICHAEL HEWITT, TERI E. KLEIN, RUSS B. ALTMAN

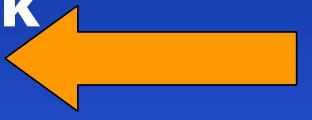
Stanford University School of Medicine
Stanford Medical Informatics

Ontology Development For A Pharmacogenetics Knowledge Base

Vortrag gehalten
am 12.06.2002
von Stefan Smers und Tilo Hielscher

Inhalt:

- **Pharmakogenetik**
- **PharmGKB**
- **Ontologie**
- **Datenmodellierung**



Crashkurs Biologie und Genetik

- **Beispiel Abfrage**
- **Beispiel Submission**

Pharmakogenetik - Problemstellung

Aufgabe:

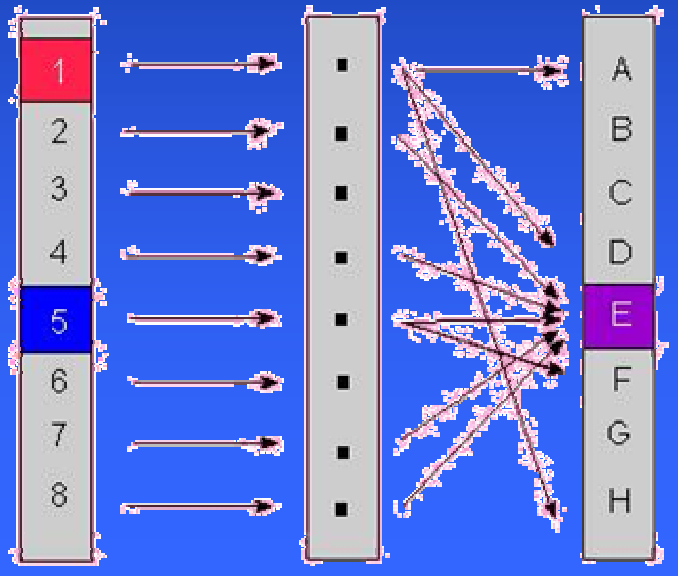
Untersuchung des Einflusses genetischer Faktoren u.a. auf die Wirksamkeit von Medikamenten

Ziel:

Erkennen einer Verbindung zwischen Genotyp und Phänotyp

Crashkurs – Genotyp vs. Phänotyp

Genotyp: **Gesamtheit aller genetisch festgelegten Merkmale und Funktionen eines Individuums**



Phänotyp: Erscheinungstyp eines Individuums, Bezeichnung für alle inneren und äußeren Strukturen und Funktionen, resultierend aus Genotyp und Umweltfaktoren;

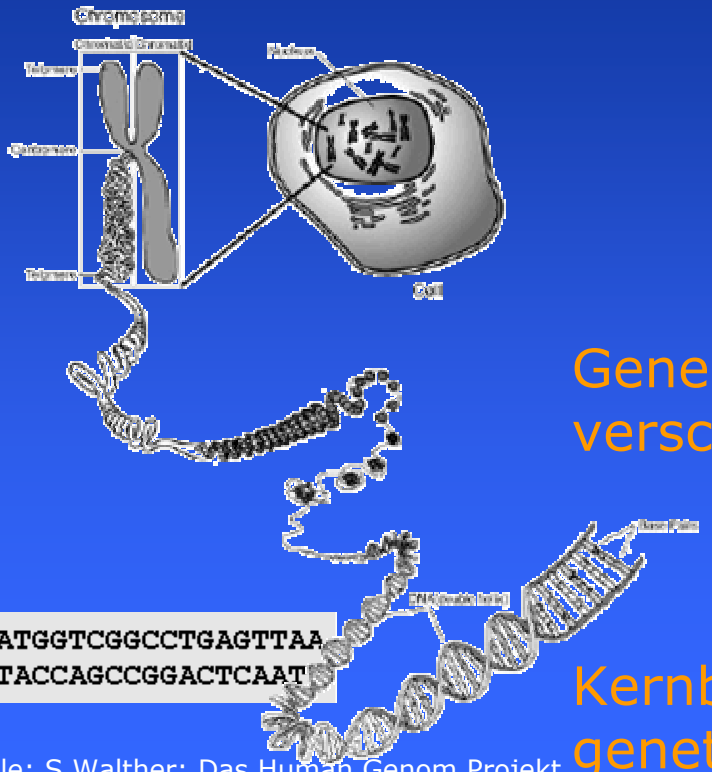
Quelle: www.botanik.uni-hamburg.de

Pharmakogenetik - Methoden

Methoden:
Daten aus Untersuchungen:
Laborexperimenten,
Berechnungen und Statistiken,
klinischen Studien

Daten z.B. Gensequenzen, Proteine,
Wirkstoffe,
messbare Ergebnisse,
Krankheiten,
funktionelle Studien

Crashkurs - Gene



Quelle: S.Walther: Das Human Genom Projekt

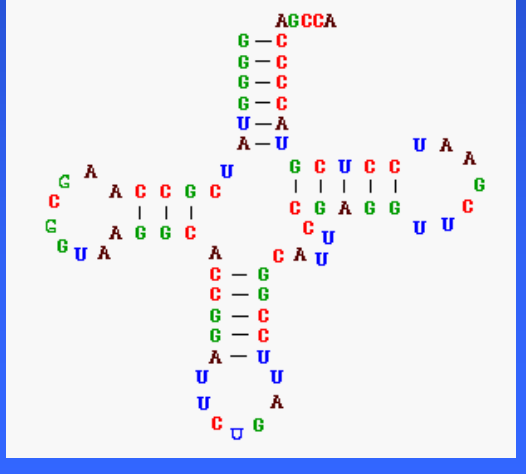
Gene bilden informations-tragende Untereinheiten des Genoms (Gesamtheit der in einer Zelle enthaltenen Erbinformationen)

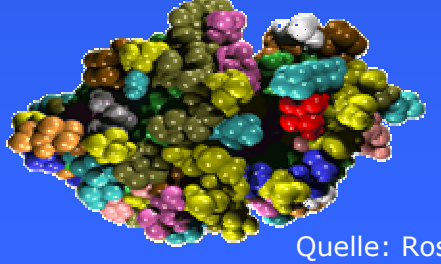
Gene sind in Nukleinsäuren verschlüsselt abgelegt. Diese enthalten die vier Kernbasen. 3-Kombinationen der Kernbasen ergeben den genetischen Code.

Crashkurs - Proteine

- Proteine

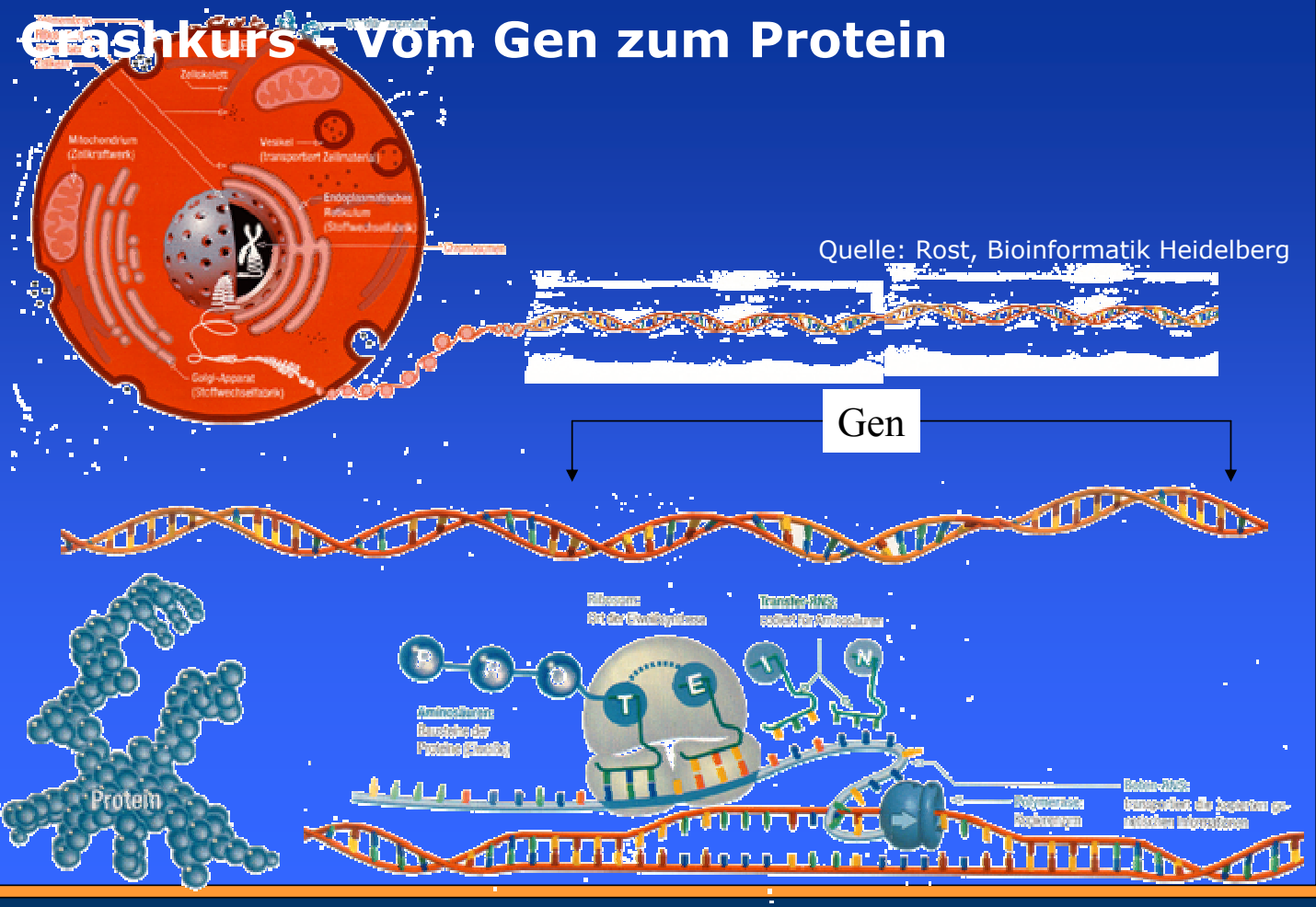
Eiweiss, Makromolekül aus vielen Aminosäuren, die über Peptidbindung miteinander verknüpft sind;
vielfältige Funktionen als Strukturelemente, Hormon oder als Katalysator biochemischer Reaktionen (Enzym)





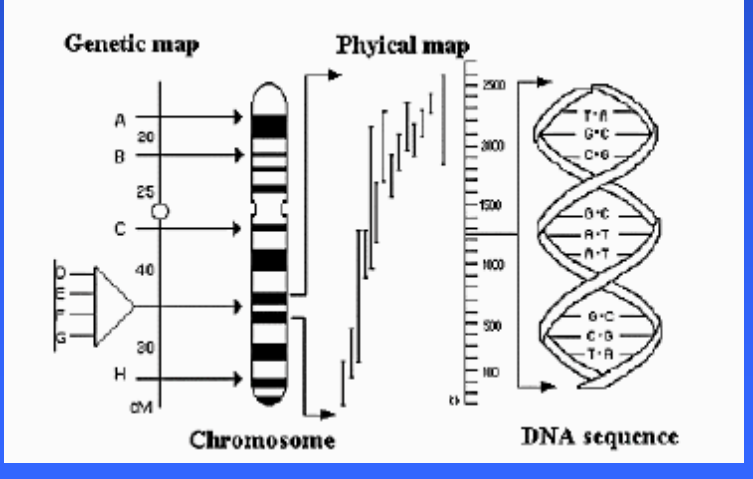
Quelle: Rost, Bioinformatik Heidelberg

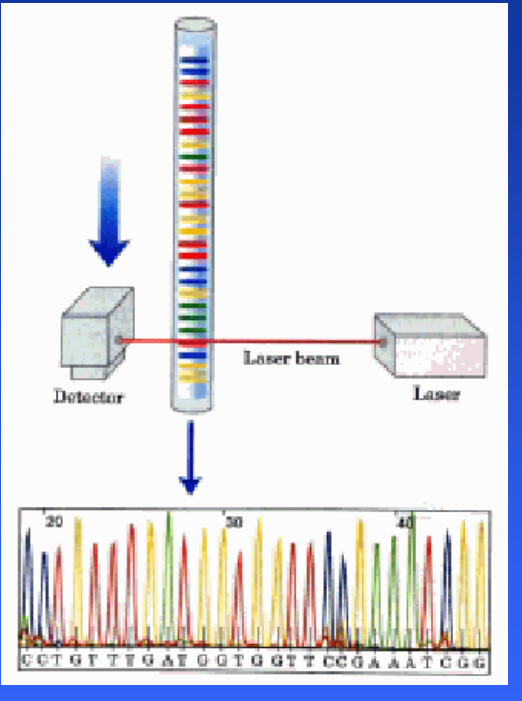
Crashkurs - Vom Gen zum Protein



Crashkurs - Methoden

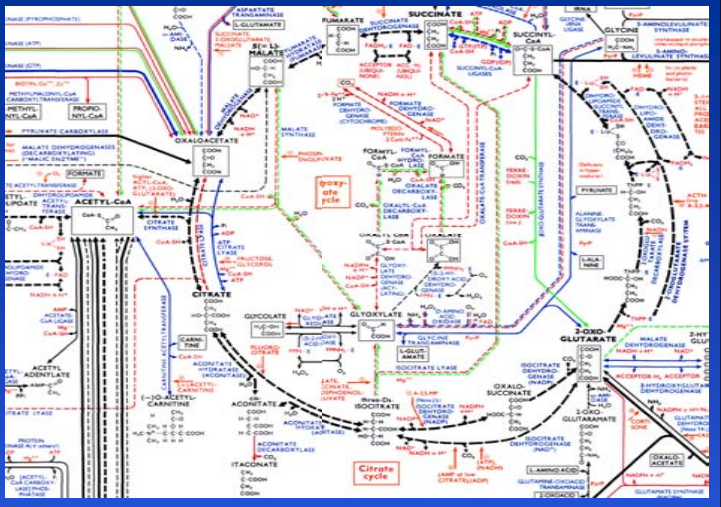
- Laborexperimente
- Berechnungen, Statistiken





Quelle: S.Walther: Das Human Genom Projekt

Pharmakogenetik - Daten



Daten sind komplex, umfangreich, vielschichtig, kommen aus verschiedenen Quellen,

Für die Forschung sind große Datenmengen erforderlich.

Die Wissenschaft ist bestrebt, gewonnene Daten zu teilen.

Quelle: Rost, Bioinformatik Heidelberg

PharmGKB – A Pharmacogenetics Knowledge Base

Mayo Clinic
(R. Weinshilboum)

University of Chicago
(M. Ratain)

University of California at San Francisco
(K. Giacomini)

University of Indiana
(D. Flockhardt)

Harvard University
(S. Weiss)

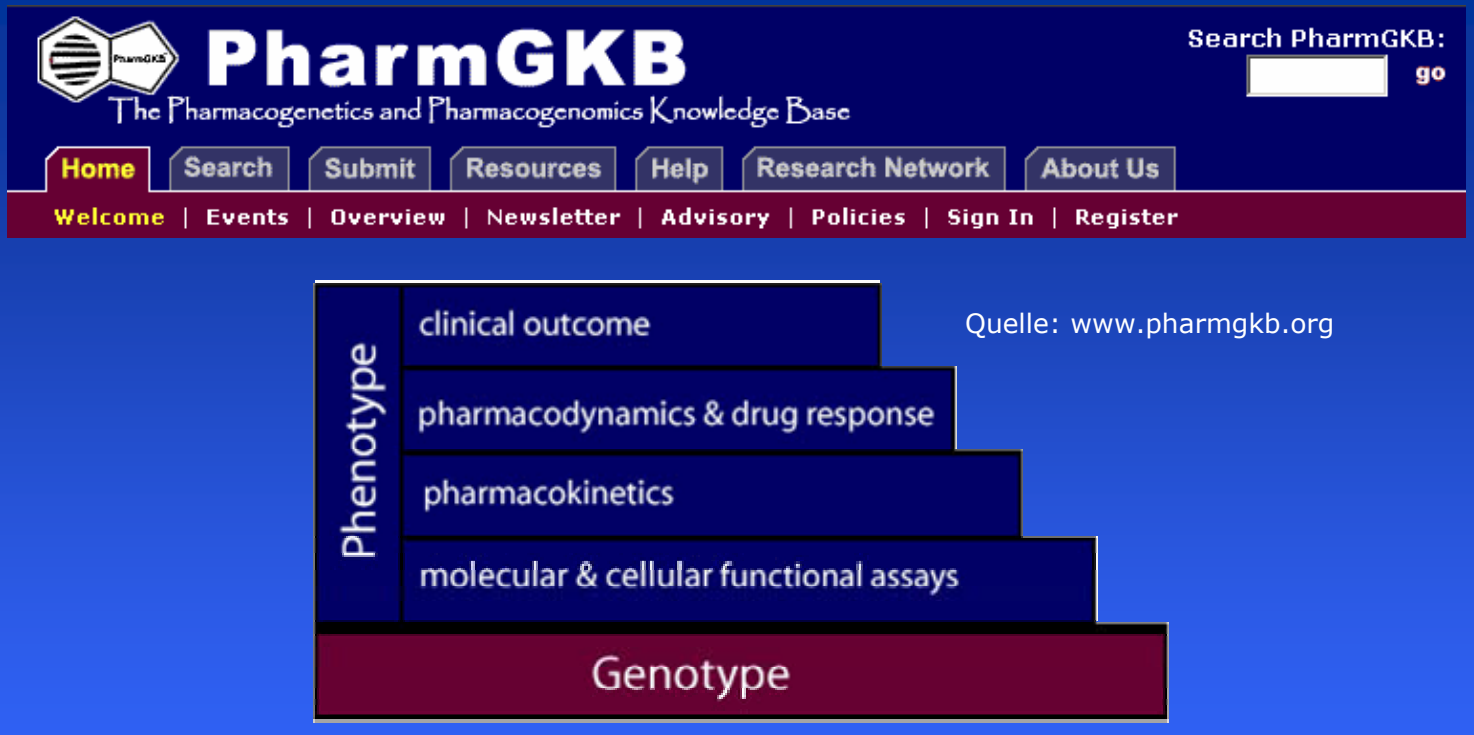
Stanford University

R. Altman
T. Klein
R. Fergerson
M. Hewett
T. Jung
S. Lin
D. Oliver
D. Rubin
M. Woon

NATIONAL INSTITUTES OF HEALTH

NIGMS
NHLBI
NHGRI
NIEHS
NCI
NLM

PharmGKB – A Pharmacogenetics Knowledge Base



Quelle: www.pharmgkb.org

Sammlung, Speicherung und Bereitstellung von Daten (data)
und Fachwissen (domain knowledge)

PharmGKB – A Pharmacogenetics Knowledge Base

PharmGKB
The Pharmacogenetics and Pharmacogenomics Knowledge Base

Search PharmGKB: go

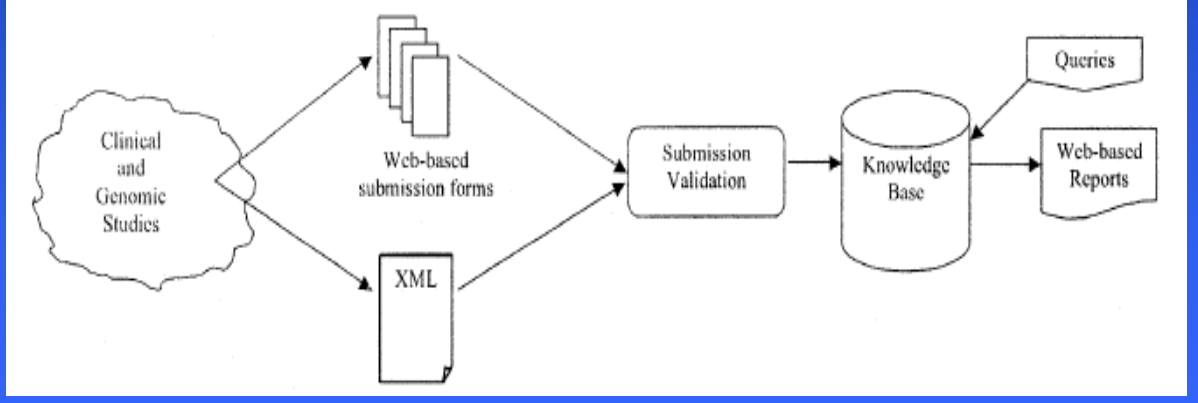
Home | Search | Submit | Resources | Help | Research Network | About Us

Welcome | Events | Overview | Newsletter | Advisory | Policies | Sign In | Register

Internet

PharmGKB

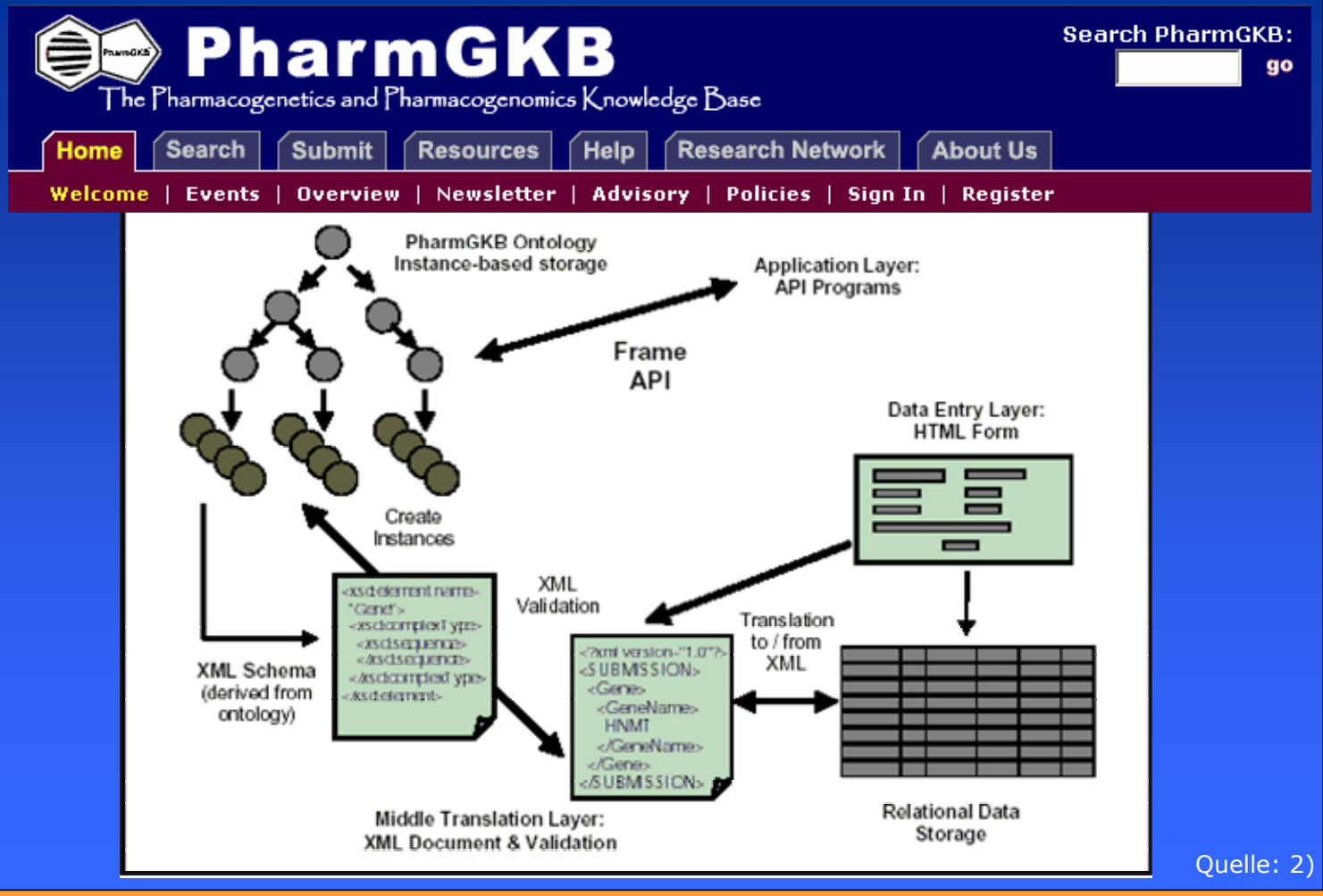
Forscher



Forscher, User

Quelle: 3)

PharmGKB – A Pharmacogenetics Knowledge Base



Quelle: 2)

PharmGKB – A Pharmacogenetics Knowledge Base


```
<ForwardPcrPrimer>
<DisplayName>Exon 5 Forward Primer</DisplayName>
<FirstAnnealingPositionInPrimer>19</FirstAnnealingPositionInPrimer>
<FirstAnnealingPositionInRegion>6</FirstAnnealingPositionInRegion>
<LastAnnealingPositionInPrimer>41</LastAnnealingPositionInPrimer>
<LastAnnealingPositionInRegion>28</LastAnnealingPositionInRegion>
<Sequence>TGTAAAACGACGGCCAGTAGGAGTATCTAGCCCAAGCAATA</Sequence>
</ForwardPcrPrimer>
```

Daten-Input im XML-Format

Quelle: 1)

Ontologie allgemein

Philosophie - Beschreibung eines Bereiches der Metaphysik;

KI-Gemeinde - Bedeutungswechsel des Begriffs

Guarino: Beschreibung eines Gegenstandes durch einen festgelegten Wortschatz, Hierarchie von Konzepten, miteinander verbunden durch Ordnungsrelationen, Anreicherung durch Axiome für Erweiterung von Relationen zwischen Konzepten oder Beschränkungen Interpretationsbreite von Konzepten

Ontologie von PharmGKB 1

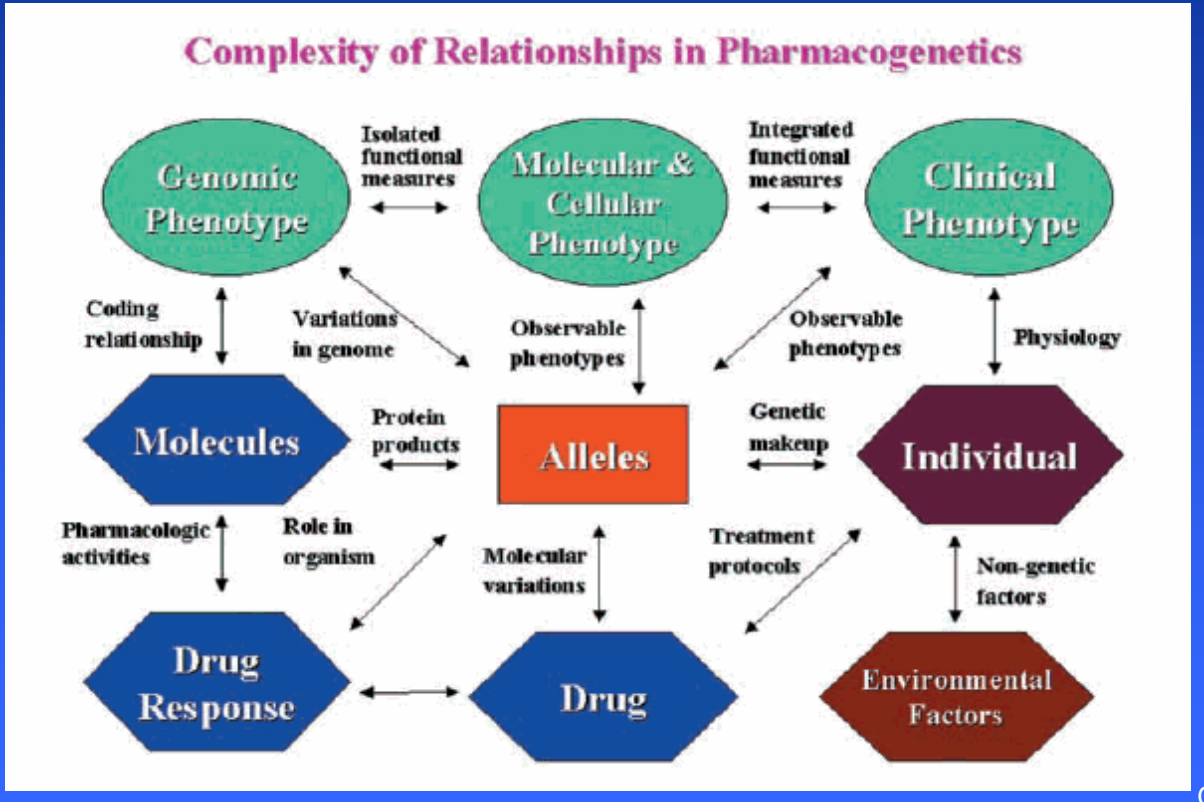
- Sammlung von Ausdrücken
- Einteilung in Hierarchien
- Standardnomenklatur

- dient zur Spezifikation von Namen und Synonymen

- Klassen und dazugehörige Slots
- Beschreibung eines bestimmten Wissensgebietes
- Schema einer Wissensbasis

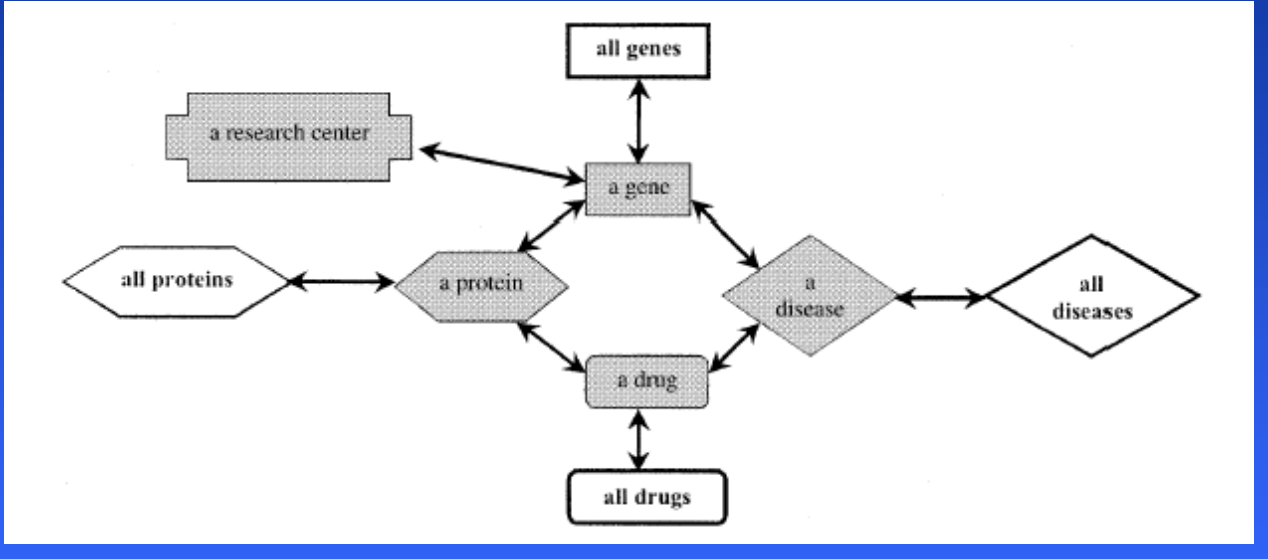
- Spezifikation Art und Ort der Speicherung von Daten und Wissen

Datenmodellierung – Komplexität pharmakogenetischer Daten



Quelle: 5)

Datenmodellierung – Darstellung der experimentellen Daten



bottom-up-Methode

Quelle: 3)

Datenmodellierung – Darstellung des Fachwissens

Kategorie	Entitäten
Gene	TPMT, HNMT, COMT, CYP2D6, UGT1A1
Proteine Enzyme Transporter Rezeptoren	 Methyl-Transferasen, Isoenzyme des Zellfarbstoffs P450 Serotonin-Transporter (SERT), Interleukin-Rezeptoren,
Wirkstoffe	inhalierte Beta-Agonisten, Topoisomerase-Hemmer,
Krankheiten	Asthma, Depression, Brustkrebs, Leukämie, Darmkrebs
Studien in vitro in vivo	 Studien der Enzymkinetik klinische Studien über Wirksamkeit und Toxizität von Wirkstoffen

Datenmodellierung – Protégé

Gestaltung
von Schnittstellen
für Wissenserwerb und Dateneingabe

- halbautomatische Generierung von Schnittstellen,
- erfolgt logisch stringent aus Wissensstrukturen (Ontologien),
- dyn. Anpassung an Änderungen der Ontologie und individueller Bedürfnisse

Herausforderungen

Vielfalt der pharmakogenetischen Daten

- Gendaten
- molekulare und zelluläre Daten
- klinische Daten

Standards

- Gennamen - Human Genome Nomenclature Committee
- Wirkstoffnamen - UMLS
- klinische Daten – keine Standards, eigener Thesaurus

Datenintegration aus verschiedenen Quellen

- komplexe Anfragen bedingen Benutzung mehrerer Datenbanken
- Datenbank-Konsolidierung
- föderierte Datenbanken

Herausforderungen

Datenschutz
- Internet
- Rekonstruktion von Identitäten
- Filter, Berechtigungen

Performance, Handhabung
- Ontologie vs. relationales Modell

PharmGKB stellt um von
ontologiebasierter Wissensbasis
auf
relationale Datenbank

Email D. Oliver, Stanford, 31.05.02

PharmGKB – A Pharmacogenetics Knowledge Base

Inhalt:

- Pharmakogenetik
- PharmGKB
- Ontologie
- Datenmodellierung

- **Beispiel Abfrage**
- **Beispiel Submission**

Beispiel Topoisomerase

PharmGKB
The Pharmacogenetics and Pharmacogenomics Knowledge Base

Search PharmGKB: go

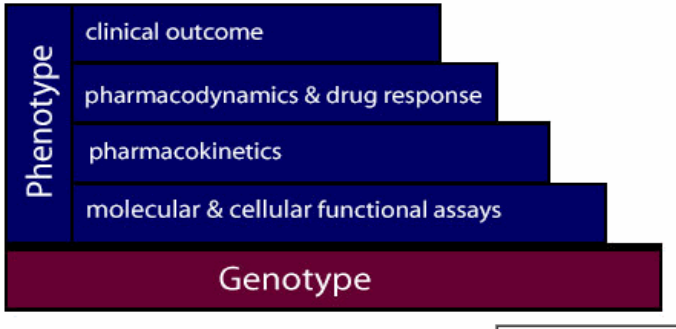
Home | Search | Submit | Resources | Help | Research Network | About Us

Welcome | Events | Overview | Newsletter | Advisory | Policies | Sign In | Register

Welcome to PharmGKB!

PharmGKB is an integrated resource about how variation in human genes leads to variation in our response to drugs. Genomic data, molecular and cellular phenotype data, and clinical phenotype data are accepted from the scientific community at large.

Categories of Pharmacogenetic Knowledge



Search the PharmGKB Knowledge Base: topoisomerase Search

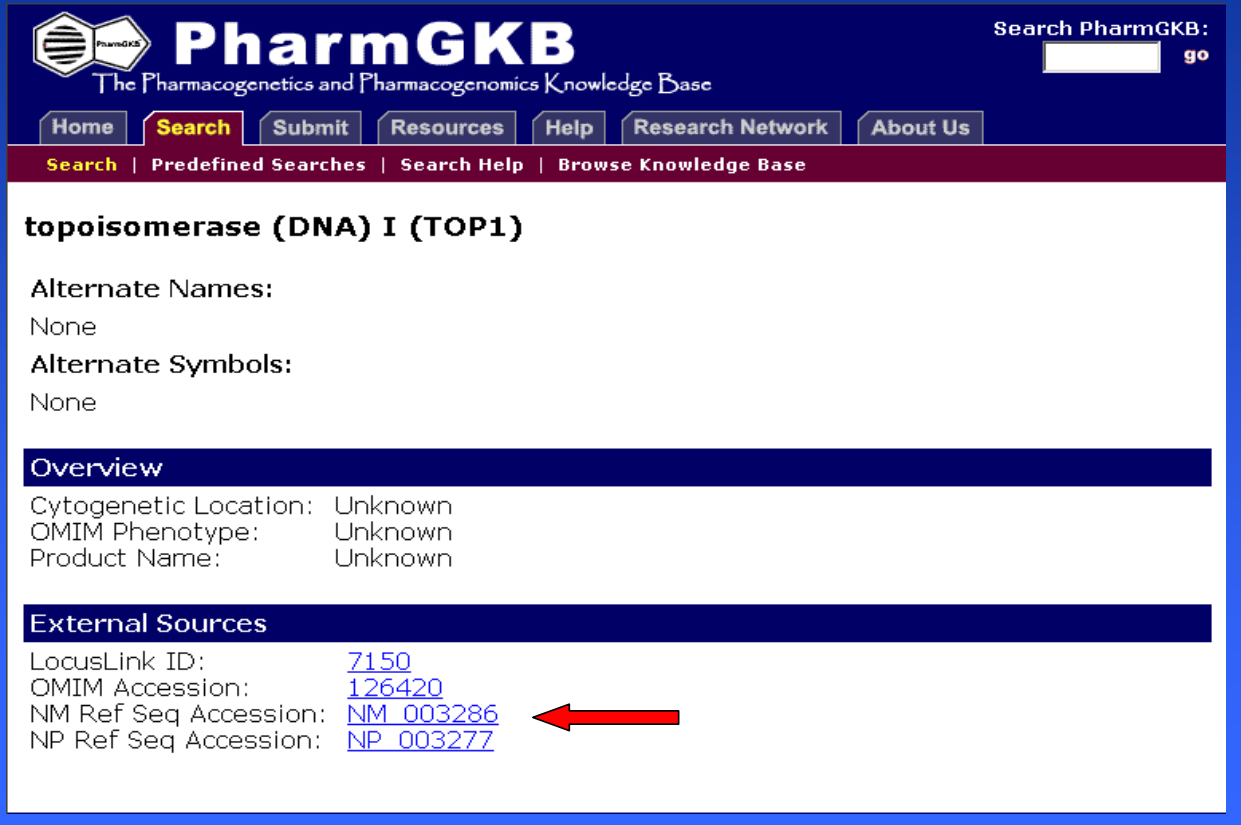
Topoisomerase ist ein Enzym. Ohne dieses können Zellen ihre Erbsubstanz nicht teilen und sterben nach einiger Zeit.

Ergebnis der Suche

Medikamente

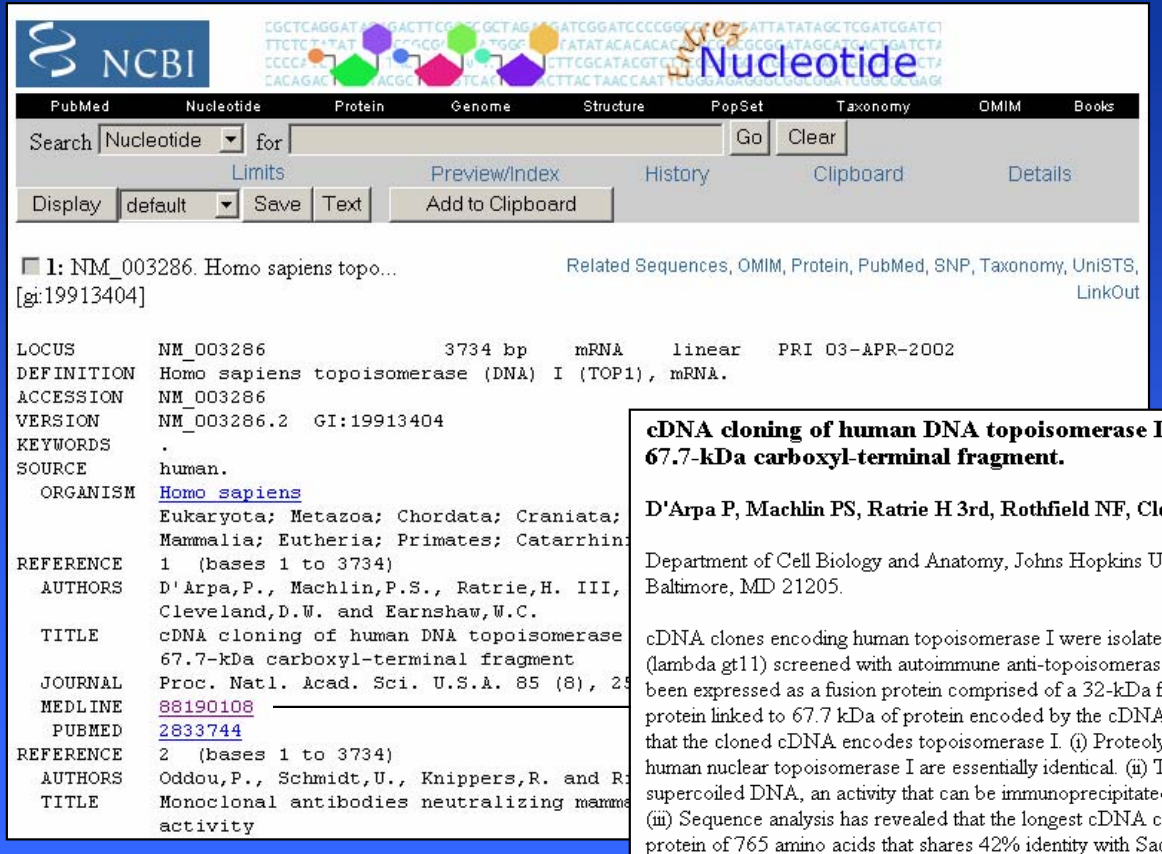
Ergebnis der Suche

Information zum Gen



NM Ref Seq Accession

National Center for Biotechnology Information



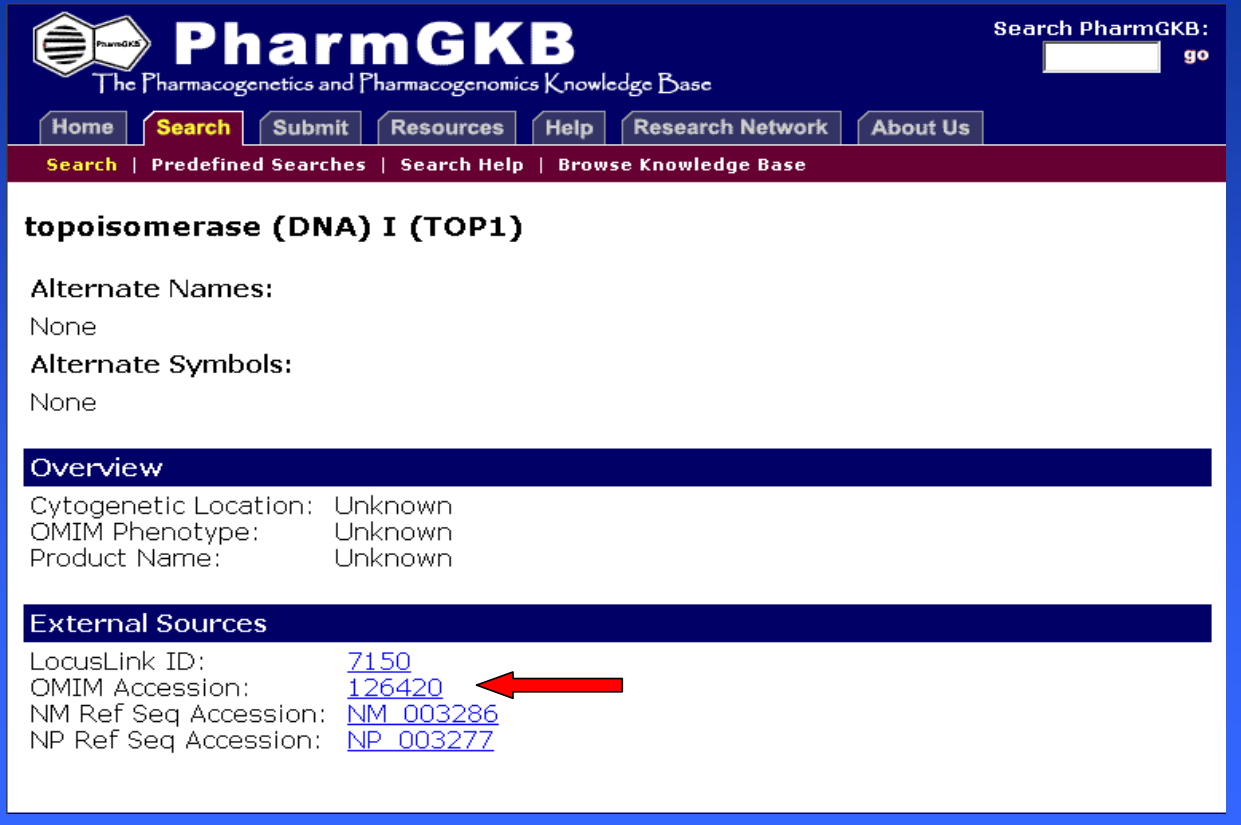
cDNA cloning of human DNA topoisomerase I: catalytic activity of a 67.7-kDa carboxyl-terminal fragment.

D'Arpa P, Machlin PS, Ratrie H 3rd, Rothfield NF, Cleveland DW, Earnshaw WC.

Department of Cell Biology and Anatomy, Johns Hopkins University School of Medicine, Baltimore, MD 21205.

cDNA clones encoding human topoisomerase I were isolated from an expression vector library (lambda gt11) screened with autoimmune anti-topoisomerase I serum. One of these clones has been expressed as a fusion protein comprised of a 32-kDa fragment of the bacterial TrpE protein linked to 67.7 kDa of protein encoded by the cDNA. Three lines of evidence indicate that the cloned cDNA encodes topoisomerase I. (i) Proteolysis maps of the fusion protein and human nuclear topoisomerase I are essentially identical. (ii) The fusion protein relaxes supercoiled DNA, an activity that can be immunoprecipitated by anti-topoisomerase I serum. (iii) Sequence analysis has revealed that the longest cDNA clone (3645 base pairs) encodes a protein of 765 amino acids that shares 42% identity with Saccharomyces cerevisiae topoisomerase I. The sequence data also show that the catalytically active 67.7-kDa fragment is

Information zum Gen



OMIM Accession

Online Mendelian Inheritance In Man

MEDLINE | Protein | DNA | LocusLink | Gene Map | GDB | MGD | Nomenclature

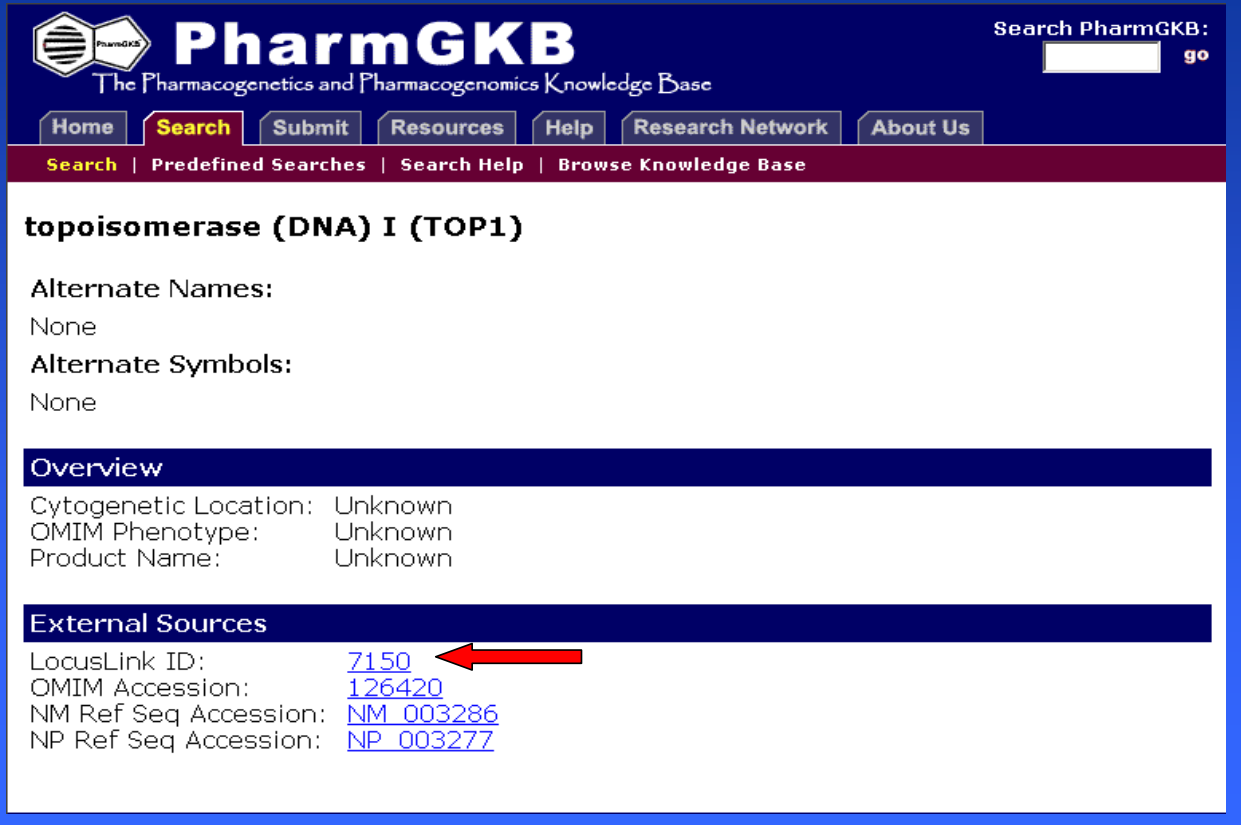
Gene Map Locus: 20q12-q13.1

Note: pressing the symbol will find the citations in MEDLINE whose text most closely matches the text of the preceding OMIM paragraph, using the Entrez MEDLINE neighboring function.

TEXT

DNA topoisomerases I and II (126430) catalyze the breaking and rejoining of DNA strands in a way that allows the strands to pass through one another, thus altering the topology of DNA. Type I topoisomerases break a single DNA strand, whereas the type II enzymes break 2 strands of duplex DNA. Several lines of evidence suggest that topoisomerase I normally functions during transcription. Whether topoisomerase I and II can substitute for each other in transcription and replication in higher eukaryotes, as they can in yeast, has not been demonstrated. D'Arpa et al. (1988) reported the cDNA cloning of human topoisomerase I. See 208900 and 227650 for disorders in which abnormality of DNA topoisomerase has been suggested. Eukaryotic DNA topoisomerase I (EC 5.99.1.2) was found in extracts of mouse cells in 1972 and thereafter was identified in all eukaryotes. The enzyme catalyzes interconversions between different topologic forms of DNA by transiently breaking DNA strands one at a time; it is therefore classified as a type I DNA topoisomerase. Juan et al. (1988) isolated cDNA clones of the human TOP1 gene by immunochemical screening of lambda phage libraries expressing human cDNA segments, using rabbit antibodies raised against purified HeLa DNA topoisomerase I. They showed that the human TOP1 is a single-copy gene. Furthermore, they mapped the gene to 20q12-q13.2 by a combination of in situ hybridization and somatic cell hybridization. By Southern blotting of digested DNA from a panel of rodent-human somatic cell hybrids, Kunze et al. (1989) demonstrated that the complete gene is located on chromosome 20 and that 2 truncated pseudogenes are located on chromosomes 1 and 22. In situ hybridization showed that the complete gene is on band 20q11.2-q13.1 and that the pseudogenes are on bands 1q23-q24 and 22q11.2-q13.1. Kunze et al. (1991) demonstrated that the coding sequence of TOP1 is split into 21 exons distributed over at least 85 kb of genomic DNA. The sizes of the 20 introns varied between 0.2 and at least 30 kb. Baumgartner et al. (1994) found a similar exon-intron structure in the Top-1 gene of the mouse and mapped it to distal chromosome 2. In addition, the mouse genome contains one truncated processed Top-1-related pseudogene on chromosome 16 where, together with the immunoglobulin lambda light chain gene, it defines a conserved linkage group common to murine chromosome 16 and human chromosome 22. The mapping data and structural features suggest that the pseudogene was established before mammalian radiation.

Information zum Gen



LocusLink

National Center for Biotechnology Information
National Library of Medicine National Institutes of Health

LocusLink

Click to Display mRNA-Genomic Alignments (spanning 95670 bps)

PUB OMIM UNIGENE MAP VAR HOMOL GDB e! UCSC

Homo sapiens Official Gene Symbol and Name (HGNC)

TOP1: topoisomerase (DNA) I

LocusID: 7150

Overview ?

RefSeq Summary: This gene encodes a DNA topoisomerase, an enzyme that controls and alters the topologic states of DNA during transcription. This enzyme catalyzes the transient breaking and rejoining of a single strand of DNA which allows the strands to pass through one another, thus altering the topology of DNA. This gene is localized to chromosome 20 and has pseudogenes which reside on chromosomes 1 and 22.

Proteome Summary: DNA topoisomerase I; relaxes supercoiled DNA

Locus Type:	gene with protein product, function known or inferred
Product:	DNA topoisomerase I
Alternate Symbols:	TOPI
Alias:	type I DNA topoisomerase

Function Submit GeneRIF (All Pubs) ?

EC Number: 5.99.1.2

GeneRIF: Gene References into Function:

11809893 • active site of a type I DNA topoisomerase from the kinetoplastid protozoan Leishmania donovani, compared to human structure

Gene Ontology™:

Term	Evidence	Source	Pub
• DNA topoisomerase I	E	Proteome	pm

Other Ontologies:

Term	Evidence	Source	Pub
• Isomerase	E	Proteome	pm
• Nuclear	NR	Proteome	pm
• Topoisomerase	E	Proteome	pm
• DNA-associated (direct or indirect)	NR	Proteome	pm

Relationships ?

Mouse Homology Maps:

NCBI vs. MGD	2 92.00 cM	Top1	Hs Mm
UCSC vs. MGD	2 92.00 cM	Top1	Hs Mm

PharmGKB – A Pharmacogenetics Knowledge Base

Map

Homo sapiens **Map View** **build 29** BLAST the Human Genome

Chromosome:
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 [**20**] 21 22 X Y

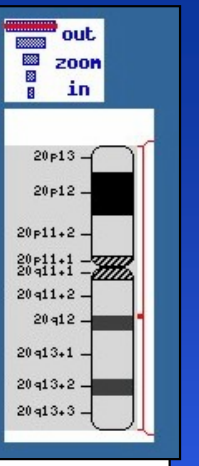
Query: TOP1 [clear]

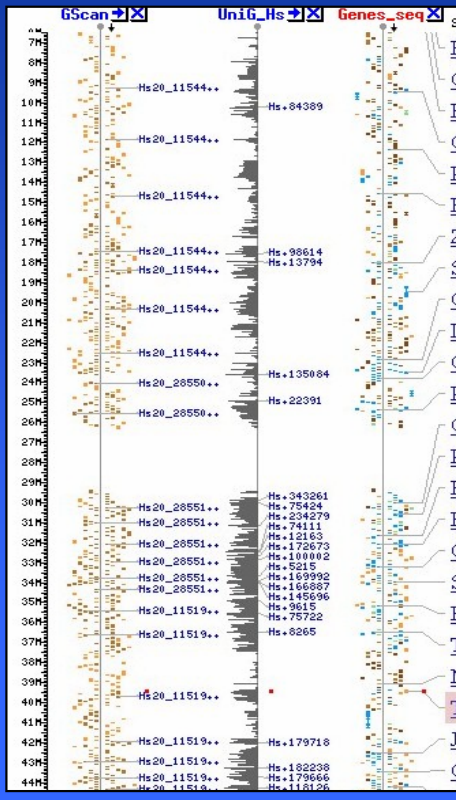
Master Map: Genes On Sequence — Maps & Options

Total Genes On Chromosome: **1179**

Region Displayed: **0-62M bp** Download/View Sequence/Evidence

Genes Labeled: **40** Total Genes in Region: **1171**

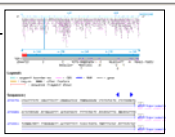
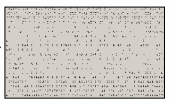
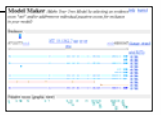




symbol	orient.	links	evidence	cyto.	full name
KIAA0552	↑	sv ev - seq mm	C	20p13	KIAA0552 gene produ
C20orf48	↓	sv ev hm seq mm	C	20p13	chromosome 20 open r
KIAA1434	↑	sv ev hm seq mm	C	20p12.3	hypothetical protein FL
C20orf103	↓	sv ev - seq mm	C	20p12	chromosome 20 open r
PA2G4P2	↓	sv ev - seq mm	C	20	proliferation-associated
RNF11B	↑	sv ev - seq mm	C	20	ring finger protein 11B,
ZNF339	↑	sv ev hm seq mm	C	20pter-q11.23	zinc finger protein 339
SLC24A3	↓	sv ev hm seq mm	C	20p13	solute carrier family 24
CYB5P4	↑	sv ev - seq mm	C	20p11.2	cytochrome b-5 pseudo
IMAGE:4837709	↓	sv ev - seq mm	C	20p11.21	cystatin pseudogene
CST2	↑	sv ev - seq mm	C	20p11.21	cystatin SA
PPIAP2	↑	sv ev - seq mm	C	20	peptidylprolyl isomerase A (cyclophilin A) ps
COX4I2	↓	sv ev hm seq mm	C	20q11.1	cytochrome c oxidase subunit IV isoform 2
POFUT1	↓	sv ev hm seq mm	C	20p11	protein O-fucosyltransferase 1
RPL12P3	↑	sv ev - seq mm	C	20q11.21	ribosomal protein L12 pseudogene 3
RPL31P2	↑	sv ev - seq mm	C	20q11.2	ribosomal protein L31 pseudogene 2
GGTL3	↑	sv ev - seq mm	C	20q11.22	gamma-glutamyltransferase-like 3
SPAG4	↓	sv ev - seq mm	C	20q11.2	sperm associated antigen 4
KIAA0889	↑	sv ev - seq mm	C	20q11.22	KIAA0889 protein
TGM2	↑	sv ev hm seq mm	C	20q12	transglutaminase 2 (C polypeptide, protein-gl
MAFB	↑	sv ev hm seq mm	C	20q11.2-q13.1	v-maf musculoaponeurotic fibrosarcoma onc
TOP1	↓	sv ev hm seq mm	?	20q12-q13.1	topoisomerase (DNA) I
JPH2	↑	sv ev - seq mm	C	20q13.11	junctophilin 2
C20orf122	↑	sv ev - seq mm	C	20q12-q13.1	chromosome 20 open reading frame 122

Legende Links

symbol	orient.	links	evidence	cyto.	full name
TOP1	↓	sv ev hm seq mm	?	20q12-q13.1	topoisomerase (DNA) I

Linked Text	Link Action	Description
Map element	Map View	The results of a genome-wide search list the map elements that contain your search term. They can be from one or more maps. Following the link for a particular map element leads to a graphical Map View of the chromosomal region that contains the element. The Master Map shown in that view will vary, reflecting the map on which a particular element was placed.
sv	Sequence Viewer 	Graphically shows the position of the map element within the sequence region. The display includes a **graphic depiction of the coding region (CDS), RNA, and gene features** that have been annotated on that sequence region. A **2 Kb section of sequence** is shown below that, with corresponding graphic annotations of the features. The **left and right arrows** at either end of the sequence data allow you to **move upstream and downstream**.
ev	Evidence Viewer	Graphical display of the biological evidence supporting a particular gene model. It displays all RefSeq models, GenBank mRNAs, annotated known or potential transcripts, and ESTs that align to the genomic sequence region of interest. (more...)
seq	Sequence Download 	Opens a form that allows you to download a region of a chromosome. The form has two parts: (1) the top part allows you to enter chromosome coordinates in text boxes, and (2) the bottom part displays the NT_* contigs (or portions of them) that are found in that chromosome region.
mm	Model Maker 	Allows you to view the evidence that was used to build a gene model on assembled genomic sequence, and to create your own version of the model by selecting exons of interest. Model Maker is accessible from sequence maps that were analyzed at NCBI and displayed in Map Viewer. To see an **example**, follow the **"mm" link** beside any gene annotated on the human "Gene_Sequence" map.
Organism Specific Links:		
fb	FlyBase 	Leads to the FlyBase Report for a map element, which includes cross-references to genome annotation data and homologs in other organisms. (more...)
hm	Human-Mouse Homology Map	a table comparing genes in homologous segments of DNA from human and mouse, sorted by position in each genome. (more...)

Farben für die Evidenz

symbol	orient.	links	evidence	cyto.	full name
TOP1	↓	sv ev hm seq mm	?	20q12-q13.1	topoisomerase (DNA) I

Gene models are shown in five colors, depending on the type of evidence that was used to construct the models. The **one or two letter code** shown in the **evidence column** (that is displayed when Gene_Sequence is the master map) also indicates the type of evidence.

Gene Color	Evidence Code	Type of evidence used to construct gene model
Blue	C	Confirmed gene model - model based on alignment of mRNA, or mRNAs plus ESTs, to the genomic sequence (see *additional notes*, below)
Light Green	E	EST only - model based on EST evidence only
Dark Brown	PE	Predicted+EST - model predicted by GenomeScan and EST evidence (more about GenomeScan)
Light Brown	P	Predicted only - model predicted by GenomeScan (more about GenomeScan)
Orange	?	Conflict - there is some discrepancy between the mRNA sequence and the gene model (see *additional notes*, below)
	I	Interim LocusID - model based alignment of mRNAs, or mRNAs plus ESTs, to the genome, in which the aligning transcripts could not be unambiguously assigned to a preexisting LocusID (see *additional notes*, below)

PharmGKB – A Pharmacogenetics Knowledge Base

Map

Homo sapiens **Map View** **build 29** BLAST the Human Genome

Chromosome:
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 [**20**] 21 22 X Y

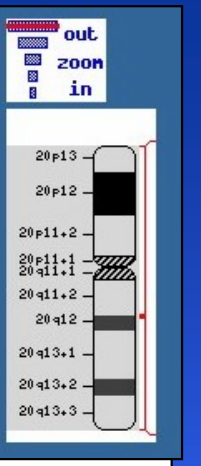
Query: TOP1 [clear]

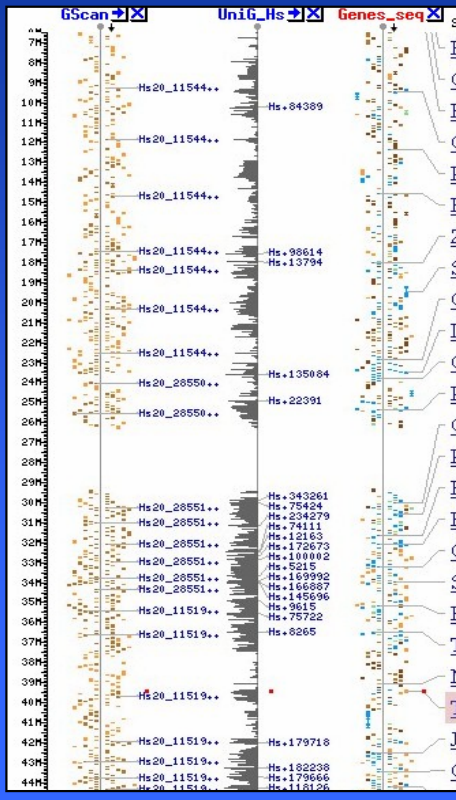
Master Map: Genes On Sequence — Maps & Options

Total Genes On Chromosome: **1179**

Region Displayed: **0-62M bp** Download/View Sequence/Evidence

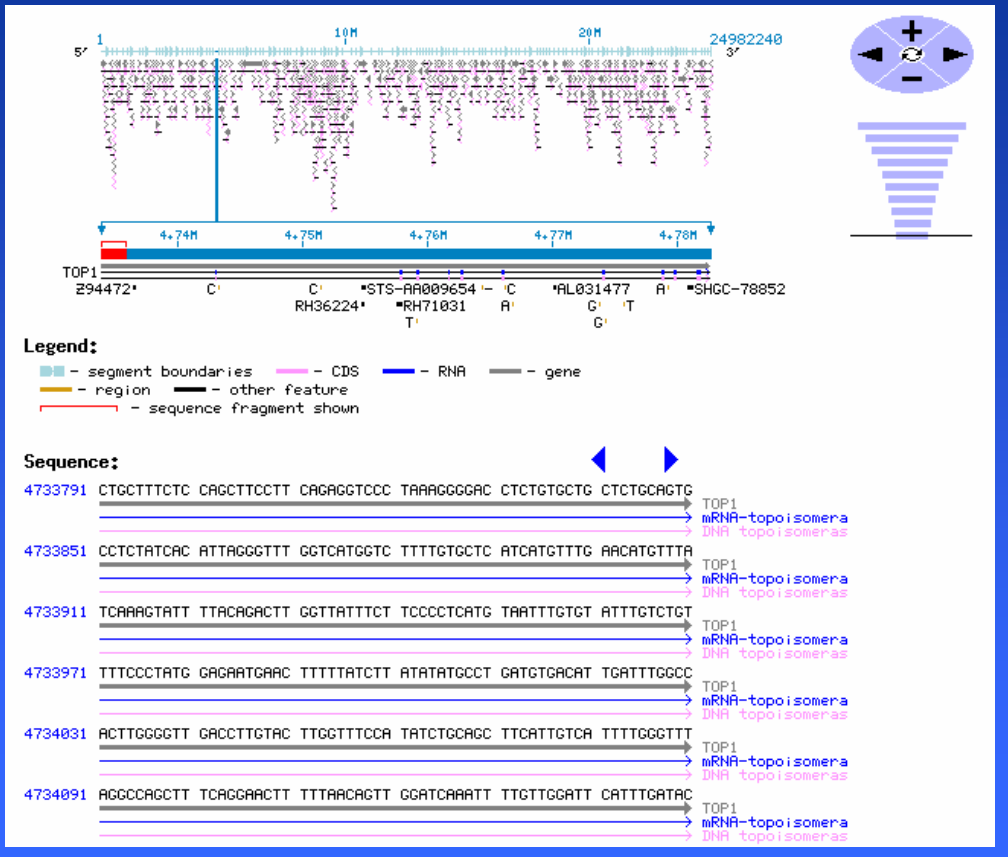
Genes Labeled: **40** Total Genes in Region: **1171**





symbol	orient.	links	evidence	cyto.	full name
KIAA0552	↑	sv ev - seq mm	C	20p13	KIAA0552 gene produ
C20orf48	↓	sv ev hm seq mm	C	20p13	chromosome 20 open r
KIAA1434	↑	sv ev hm seq mm	C	20p12.3	hypothetical protein FL
C20orf103	↓	sv ev - seq mm	C	20p12	chromosome 20 open r
PA2G4P2	↓	sv ev - seq mm	C	20	proliferation-associated
RNF11B	↑	sv ev - seq mm	C	20	ring finger protein 11B,
ZNF339	↑	sv ev hm seq mm	C	20pter-q11.23	zinc finger protein 339
SLC24A3	↓	sv ev hm seq mm	C	20p13	solute carrier family 24
CYB5P4	↑	sv ev - seq mm	C	20p11.2	cytochrome b-5 pseudo
IMAGE:4837709	↓	sv ev - seq mm	C	20p11.21	cystatin pseudogene
CST2	↑	sv ev - seq mm	C	20p11.21	cystatin SA
PPIAP2	↑	sv ev - seq mm	C	20	peptidylprolyl isomerase A (cyclophilin A) ps
COX4I2	↓	sv ev hm seq mm	C	20q11.1	cytochrome c oxidase subunit IV isoform 2
POFUT1	↓	sv ev hm seq mm	C	20p11	protein O-fucosyltransferase 1
RPL12P3	↑	sv ev - seq mm	C	20q11.21	ribosomal protein L12 pseudogene 3
RPL31P2	↑	sv ev - seq mm	C	20q11.2	ribosomal protein L31 pseudogene 2
GGTL3	↑	sv ev - seq mm	C	20q11.22	gamma-glutamyltransferase-like 3
SPAG4	↓	sv ev - seq mm	C	20q11.2	sperm associated antigen 4
KIAA0889	↑	sv ev - seq mm	C	20q11.22	KIAA0889 protein
TGM2	↑	sv ev hm seq mm	C	20q12	transglutaminase 2 (C polypeptide, protein-g
MAFB	↑	sv ev hm seq mm	C	20q11.2-q13.1	v-maf musculoaponeurotic fibrosarcoma onc
TOP1	↓	sv ev hm seq mm	?	20q12-q13.1	topoisomerase (DNA) I
JPH2	↑	sv ev - seq mm	C	20q13.11	junctophilin 2
C20orf122	↑	sv ev - seq mm	C	20q12-q13.1	chromosome 20 open reading frame 122

Sequence View



PharmGKB – A Pharmacogenetics Knowledge Base

Chr. 20

Human Genome Resources

Select a Chromosome:

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	**20**
21	22	X	Y

MapViewer Chr 20

LocusLink Chr.20 Loci

UniGene Chr.20 Clusters

Mouse Homologies for Human Chr.20

Chromosome 20

heterochromatin
draft sequence

click on finished sequence to view the contig components

> 1000 kb
250 – 1000 kb
< 250 kb

STS Maps

Map	STSs	Type	Ref
GeneMap'99	758	RH	1
Whitehead	386	YAC	2
Stanford	258	RH	3
Genethon	144	Link	4

Sequencing Progress

Euchromatic size:	72000 kb
Finished sequence:	59424 kb
Percent finished:	82.5 %
Number of contigs:	7

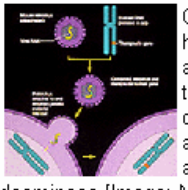
List contigs by size
List contigs by position

Sequencing is being coordinated by: The Sanger Center

Cytogenetics

Mitelman et al.(5) have cataloged recurrent chromosome aberrations associated with cancers.
Chromosome 20 aberrations

Disease Genes



Gene therapy has been attempted to treat SCID caused by the absence of adenosine deaminase [Image: National Cancer Institute]

Some diseases associated with chromosome 20 mutations

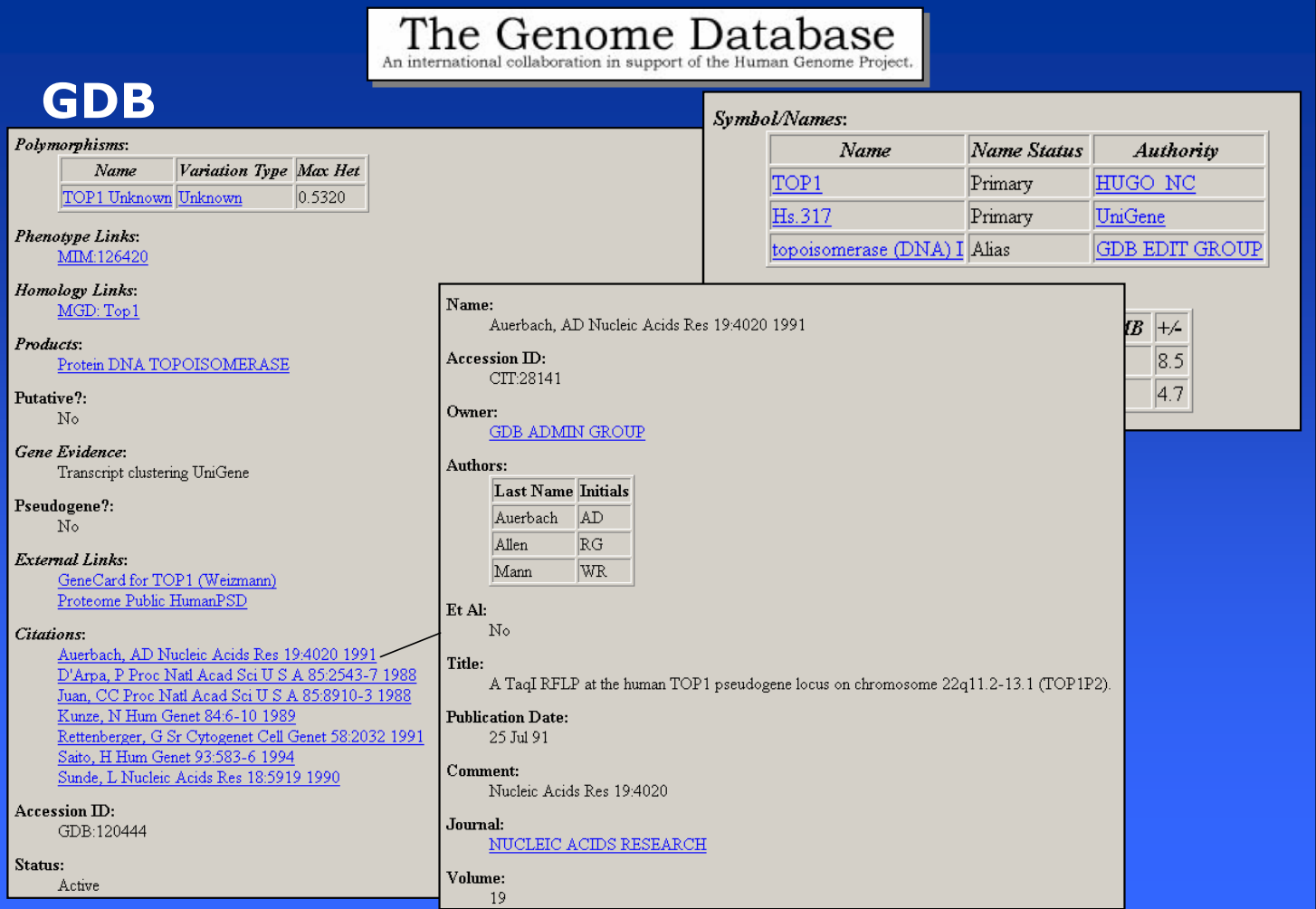
Alagille Syndrome
Corneal Dystrophy, polymorphous
SCID, due to ADA deficiency

See OMIM or Genes and Disease for additional examples and detailed information.
OMIM Home Page
OMIM Morbid Map
Genes and Disease

References

1 Deloukas et al. (1998). A physical map of 30,000 human genes. Science 282, 744-746.
2 Hudson et al. (1995). An STS-based map of the human genome. Science 270, 1945-1954.
3 Stewart et al. (1997). An STS-based radiation hybrid map of the human genome. Genome Res 7, 422-433.
4 Dib, et al. (1995). A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380, 152-154.
5 Mitelman, et al. (1997). A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. Nat Genet Apr 15 Suppl., 417-474.

PharmGKB – A Pharmacogenetics Knowledge Base



PubMed

PubMed National Library of Medicine NLM

1: Das A, Mandal C, Dasgupta A, Sengupta T, Majumder HK.
An insight into the active site of a type I DNA topoisomerase from
donovani.
Nucleic Acids Res. 2002 Feb 1;30(3):794-802.
PMID: 11809893 [PubMed - indexed for MEDLINE]

2: Ireton GC, Stewart L, Parker LH, Champoux JJ.
Expression of human topoisomerase I with a partial deletion of the
enzymes that respond differently to camptothecin.
J Biol Chem. 2000 Aug 18;275(33):25820-30.
PMID: 10827183 [PubMed - indexed for MEDLINE]

3: Redinbo MR, Stewart L, Kuhn P, Champoux JJ, Hol WG.
Crystal structures of human topoisomerase I in covalent and nonco
Science. 1998 Mar 6;279(5356):1504-13.
PMID: 9488644 [PubMed - indexed for MEDLINE]

4: Friedberg EC.
Relationships between DNA repair and transcription.
Annu Rev Biochem. 1996;65:15-42. Review.
PMID: 8811173 [PubMed - indexed for MEDLINE]

5: Fujimori A, Harker WG, Kohlhagen G, Hoki Y, Pommier Y.
Mutation at the catalytic site of topoisomerase I in CEM/C2, a hun
camptothecin.
Cancer Res. 1995 Mar 15;55(6):1339-46.
PMID: 7882333 [PubMed - indexed for MEDLINE]

6: Juan CC, Hwang JL, Liu AA, Whang-Peng J, Knutsen T, Huebner K, Croce CM, Zhang H, Wang JC, Liu LF. Related Articles, Nucleotide, OMIM, Protein
Human DNA topoisomerase I is encoded by a single-copy gene that maps to chromosome region 20q12-13.2.
Proc Natl Acad Sci U S A. 1988 Dec;85(23):8910-3.
PMID: 2848244 [PubMed - indexed for MEDLINE]

1: Nucleic Acids Res 2002 Feb 1;30(3):794-802 Related Articles, Books, LinkOut

Full text article at nar.oupjournals.org

An insight into the active site of a type I DNA topoisomerase from the kinetoplastid protozoan Leishmania donovani.

Das A, Mandal C, Dasgupta A, Sengupta T, Majumder HK.

Molecular Parasitology Laboratory and Drug Design, Development and Molecular Modelling, Indian Institute of Chemical Biology, 4, Raja S.C. Mullick Road, Calcutta 700032, India.

DNA topoisomerases are ubiquitous enzymes that govern the topological interconversions of DNA thereby playing a key role in many aspects of nucleic acid metabolism. Recently determined crystal structures of topoisomerase fragments, representing nearly all the known subclasses, have been solved. The type IB enzymes are structurally distinct from other known topoisomerases but are similar to a class of enzymes referred to as tyrosine recombinases. A putative topoisomerase I open reading frame from the kinetoplastid Leishmania donovani was reported which shared a substantial degree of homology with type IB topoisomerases but having a variable C-terminus. Here we present a molecular model of the above parasite gene product, using the human topoisomerase I crystal structure in complex with a 22 bp oligonucleotide as a template. Our studies indicate that the overall structure of the parasite protein is similar to the human enzyme; however, major differences occur in the C-terminal loop, which harbors a serine in place of the usual catalytic tyrosine. Most other structural themes common to type IB topoisomerases, including secondary structural folds, hinged clamps that open and close to bind DNA, nucleophilic attack on the scissile DNA strand and formation of a ternary complex with the topoisomerase I inhibitor camptothecin could be visualized in our homology model. The validity of serine acting as the nucleophile in the case of the parasite protein model was corroborated with our biochemical mapping of the active site with topoisomerase I enzyme purified from L. donovani promastigotes.

PMID: 11809893 [PubMed - indexed for MEDLINE]

UniGene

Mensch
Maus
Ratte
Kresse
Fadenwurm
Fliege
Hefe

UniGene Cluster Hs.317 *Homo sapiens*

TOP1 Topoisomerase (DNA) I

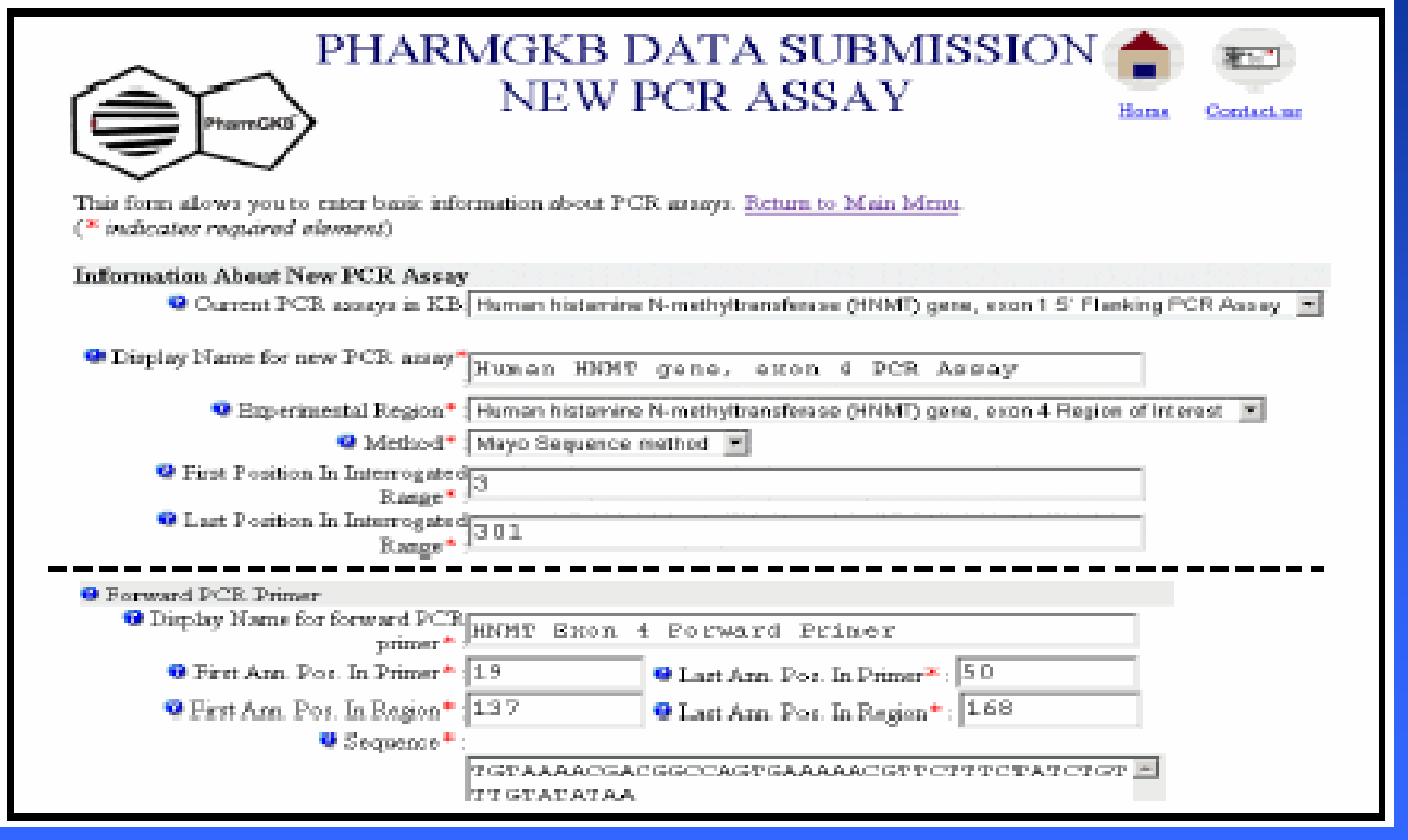
SEE ALSO
LocusLink: 7150
OMIM: 126420
HomoloGene: Hs.317

SELECTED MODEL ORGANISM PROTEIN SIMILARITIES
organism, protein and percent identity and length of aligned region

H.sapiens:	sp:P11387 - TOP1_HUMAN DNA TOPOISOMERASE I	**100 % / 764 aa** (see ProtEST)
M.musculus:	pir:JU0144 - DNA TOPOISOMERASE I	**96 % / 764 aa** (see ProtEST)
R.norvegicus:	pir:A56577 - A56577 microtubule-associated protein MAP 1B - rat	**29 % / 217 aa** (see ProtEST)
A.thaliana:	pir:S22864 - DNA TOPOISOMERASE I	**36 % / 829 aa** (see ProtEST)
C.elegans:	pir:T23648 - T23648 hypothetical protein M01E5.5a - Caenorhabditis elegans	**54 % / 731 aa** (see ProtEST)
D.melanogaster:	pir:S35521 - DNA TOPOISOMERASE I	**59 % / 700 aa** (see ProtEST)
S.cerevisiae:	pir:ISBYT1 - ISBYT1 DNA topoisomerase (EC 5.99.1.2) - yeast (Saccharomyces cerevisiae)	**42 % / 694 aa** (see ProtEST)

MAPPING INFORMATION
Chromosome: 20
OMIM Gene Map: 20q12-q13.1
Whitehead map: WI-7849, Chr.20, YAC contig WC20.4
UniSTS entries: stSG10970 Genomic Context: Map View
UniSTS entries: WI-18785 Genomic Context: Map View
UniSTS entries: WI-7849 Genomic Context: Map View

Submission



PHARMGKB DATA SUBMISSION
NEW PCR ASSAY

Home Contact us

This form allows you to enter basic information about PCR assays. Return to Main Menu
(* indicates required element)

Information About New PCR Assay

Current PCR assays in KB: Human histamine N-methyltransferase (HNMT) gene, exon 1 5' Flanking PCR Assay

Display Name for new PCR assay*: Human HNMT gene, exon 4 PCR Assay

Experimental Region*: Human histamine N-methyltransferase (HNMT) gene, exon 4 Region of Interest

Method*: Mayo Sequence method

First Position In Interrogated Range*: 3

Last Position In Interrogated Range*: 301

Forward PCR Primer

Display Name for forward PCR primer*: HNMT Exon 4 Forward Primer

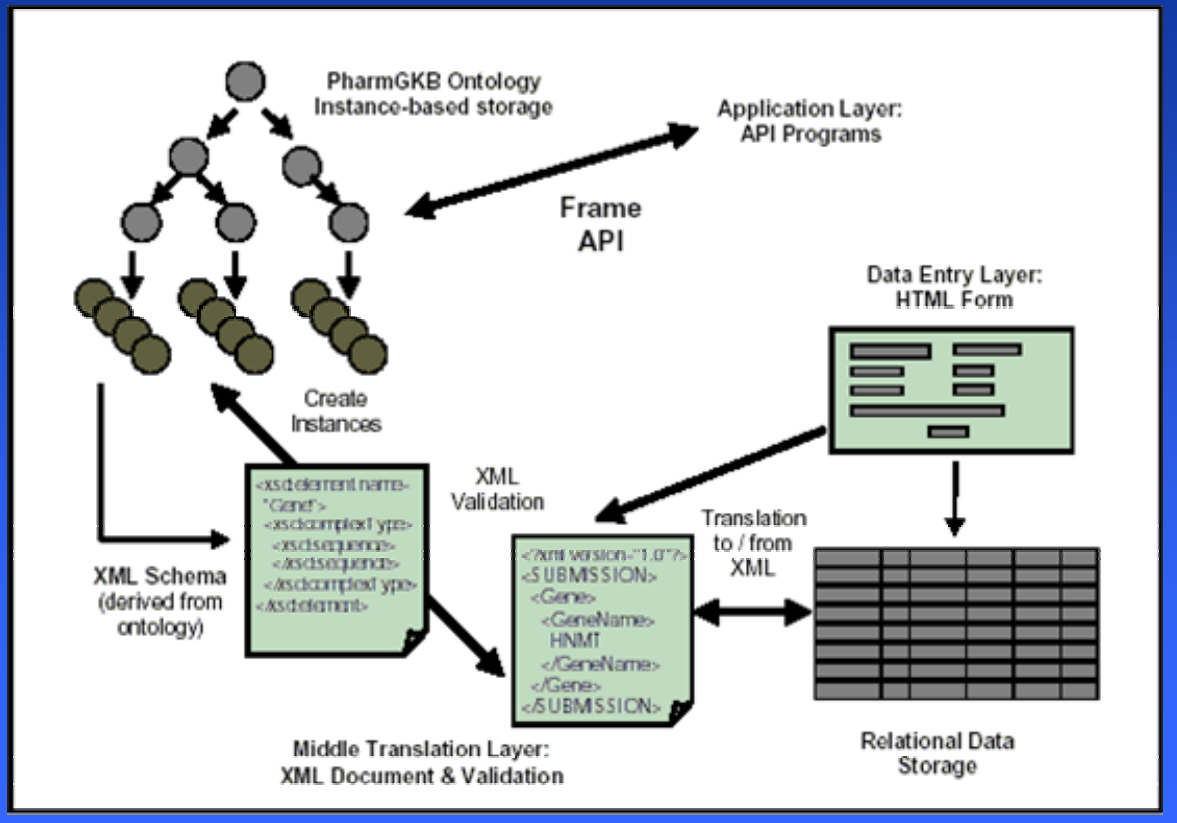
First Ann. Pos. In Primer*: 19 Last Ann. Pos. In Primer*: 50

First Ann. Pos. In Region*: 137 Last Ann. Pos. In Region*: 168

Sequence*:

TGTAAAACGACGGCCAGTGAAAAACGTTCTTTCTATCTGT
TTGTATATAA

Submission Schema



Quelle: 3)

Quellen

- Oliver et. al.: „Ontology Development for a Pharmacogenetic Knowledge Base“ [1)]
- Rubin et. al.: „Automatic Data Acquisition into Ontologies...“ [2)]
- Hewett et. al.: „PharmGKB: The Pharmacogenetic Knowledge Base“ [3)]
- Altman/Klein: „Challenges for Biomedical Informatics and Pharmacogenomics“ [4)]
- Klein et. al.: „Integrating genotype and phenotype: an overview of the PharmGKB project“ [5)]

- http://pharmgkb.org

- Stryer: Biochemistry