NCBI FieldGuide

NCBI Molecular Biology Resources

A Field Guide

Part 1

March 2, 2006 University of Medicine and Dentistry of New Jersey

NCBI FieldGuide

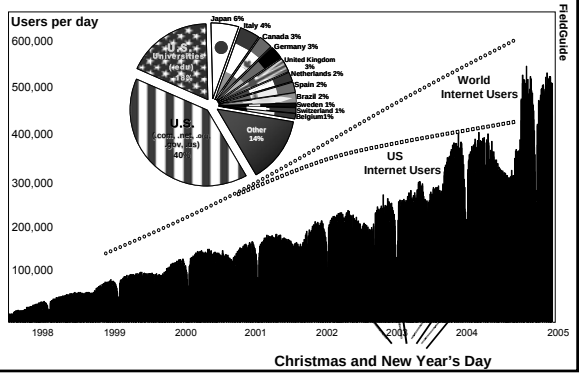
The National Center for Biotechnology Information

Created in 1988 as a part of the National Library of Medicine at NIH

- Establish public databases
- Research in computational biology
- Develop software tools for sequence analysis
- Disseminate biomedical information

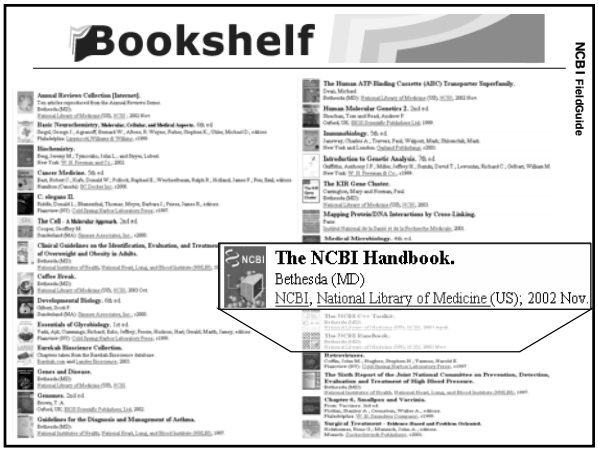
NCBI FieldGuide

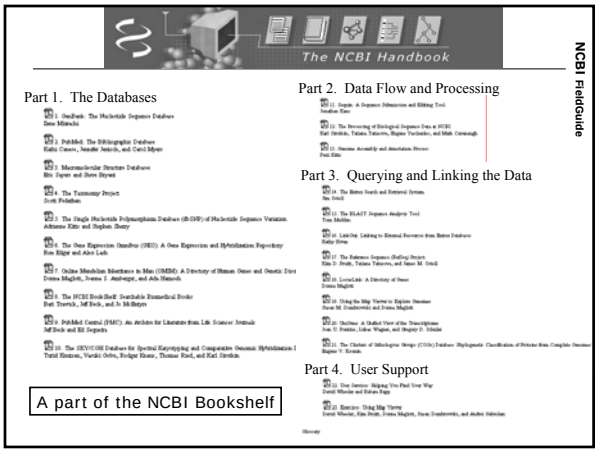
NCBI Web Traffic



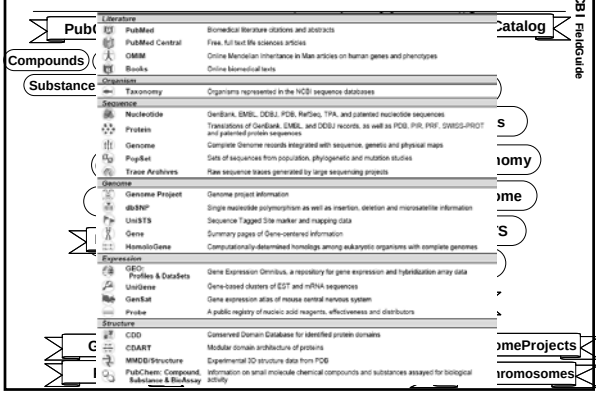
NCBI FieldGuide

Literature Databases
NCBI FieldGuide
PubMed
National Library of Medicine
PubMed Central
An archive of life sciences journals
Journals
NLM Catalog
National Library of Medicine
Bookshelf
OMIM
Online Mendelian Inheritance in Man
Johns Hopkins University


Bookshelf
NCBI FieldGuide
The NCBI Handbook.
Bethesda (MD)
NCBI, National Library of Medicine (US); 2002 Nov.


The NCBI Handbook
NCBI FieldGuide
Part 1. The Databases
Part 2. Data Flow and Processing
Part 3. Querying and Linking the Data
Part 4. User Support
A part of the NCBI Bookshelf

The *(ever expanding)* Entrez System



NCBI FieldGuide

What is Entrez?

NCBI FieldGuide

- A system of 29 *linked* databases
- A tool for finding biologically linked data
- A text search and retrieval engine
- A virtual workspace for manipulating large datasets

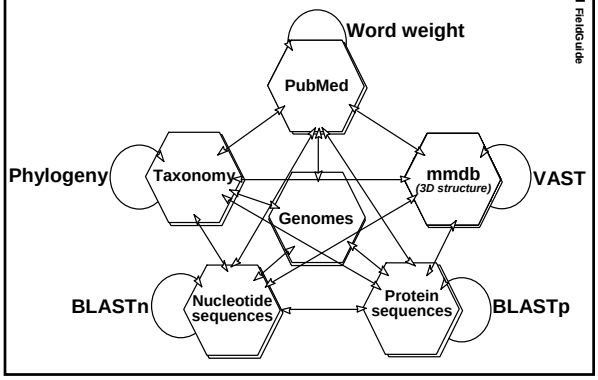
Entrez Databases

NCBI FieldGuide

- Each record is assigned a UID.
  - A “unique integer identifier” for internal tracking
- All Molecular Database entries are organized by organism ***(Taxonomy Database)***.
- Each record is indexed by data fields.
  - [author], [title], [organism], and many others
- Each record is given a Document Summary.
  - a summary of the record’s content (DocSum)
- Each record is assigned links to biologically related UIDs.

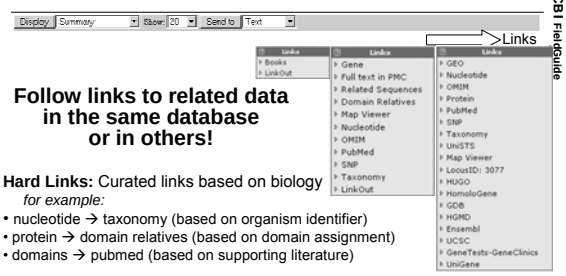
Examples of Database Integration at NCBI

NCBI FieldGuide



Following Links

NCBI FieldGuide



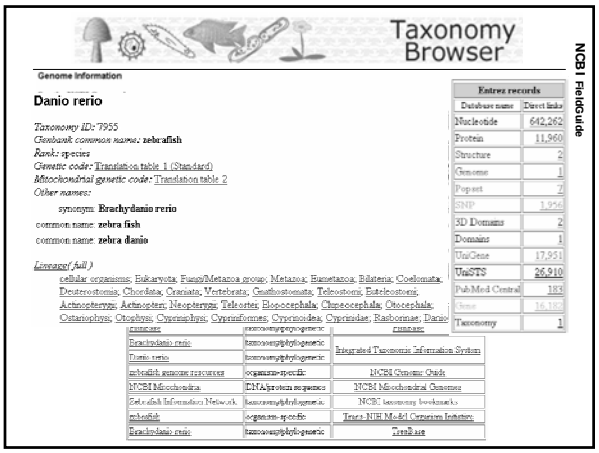
Follow links to related data in the same database or in others!

Hard Links: Curated links based on biology
for example:
- nucleotide → taxonomy (based on organism identifier)
- protein → domain relatives (based on domain assignment)
- domains → pubmed (based on supporting literature)

Soft Links: Pre-computed analyses
for example:
- nucleotide → related sequences (BLAST neighbors)
- protein → conserved domains (CDD/RPS-BLAST search)
- gene → map viewer (map position of annotated gene)

Taxonomy Browser

NCBI FieldGuide



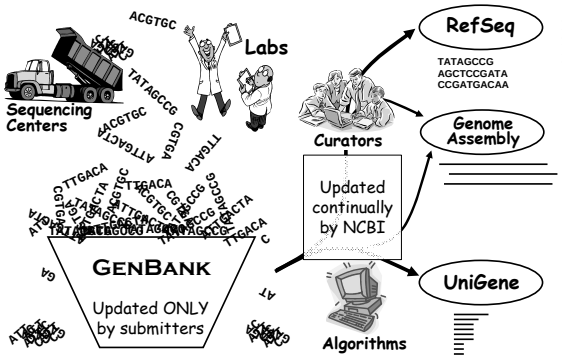
Types of Molecular Databases

NCBI FieldGuide

- **Primary Databases**
  - Raw and redundant Data…..submitted, "owned" and updated by experimentalists
    - Examples: GenBank, SNP, GEO, PubChem Substance & BioAssay
- **Derivative Databases**
  - *Human-curated* (compilation and curation of data)
    - Examples: GEO Datasets, Structure & Literature databases
  - *Computationally-Derived*
    - Example: UniGene, HomoloGene, PubChem Compound
  - *Combination*
    - Examples: RefSeq, Gene, Genome Assembly, Conserved Domain and Structure databases

Primary vs. Derivative Sequence Databases

NCBI FieldGuide



1º Sequence Database

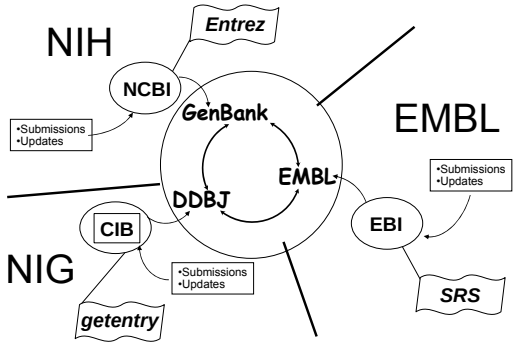
NCBI FieldGuide

GenBank

- Nucleotide only sequence database
- Archival in nature
- Each record is assigned a stable accession number
- Submission of GenBank Data to NCBI
  - **Direct submissions of individual records via Web** (BankIt, Sequin)
  - **Batch submissions of bulk sequences via Email** (EST, GSS, STS)
  - **FTP accounts for Sequencing Centers**
- Three collaborating databases and other sources of data

The International Sequence Database Collaboration

NCBI FieldGuide



GenBank

NCBI FieldGuide

Release 152	February 2006
54,584,635	Records
59,750,386,305	Nucleotides
>140,000	Species
216 Gigabytes	*908 files*

- full release every two months
- incremental and cumulative updates daily
- available only through internet

ftp://ftp.ncbi.nih.gov/genbank/

GenBank Divisions

NCBI FieldGuide

PRI (29)	Primate
ROD (23)	Rodent
PLN (17)	Plant and Fungal
BCT (13)	Bacterial/Archeal
VRT (10)	Other Vertebrate
INV (9)	Invertebrate
VRL (5)	Viral
MAM (2)	Mammalian
PHG (1)	Phage
SYN (1)	Synthetic
UNA (1)	Unannotated

EST (487)	Expressed Sequence Tag
GSS (172)	Genome Survey Sequence
HTG (79)	High Throughput Genomic
PAT (20)	Patent sequences
STS (14)	Sequence Tagged Site
HTC (10)	High Throughput cDNA
ENV (3)	Environmental Samples

Traditional

- Direct Submissions (Sequin/Bankit)
- Accurate (~1 error per 10,000 bp)
- Well characterized
- Organized by taxonomy

Bulk

- From sequencing projects
- Batch submissions (ftp/email)
- Inaccurate
- Poorly Characterized
- Organized by sequence type

Derivative Sequence Database

NCBI FieldGuide

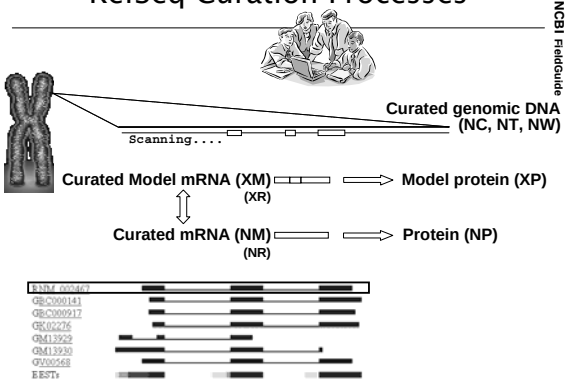
RefSeq
genomes·transcripts·proteins

RELEASE 15 IS NOW AVAILABLE ON THE FTP SITE!

- The curated "best representative" sequences
- Standardized nomenclature and record structure
- Added annotation (references, sequence features)

RefSeq Curation Processes

NCBI FieldGuide



Curated RefSeq Records

NCBI FieldGuide

```
LOCUS       ADSS      1368 bp mRNA   linear      PRI 27-AUG-2002
DEFINITION  Homo sapiens adenylosuccinate synthase (ADSS), mRNA.
ACCESSION   NM_001126
VERSION     NM_001126.1 GI:4557270
```

RefSeq Nucleotide

```
LOCUS       ADSS       455 aa         linear      PRI 27-AUG-2002
DEFINITION  adenylosuccinate synthase; Adenylosuccinate synthetase
(Ade(-)H-complementing) Homo sapiens .
ACCESSION   NP_001117
VERSION     NP_001117.1 GI:4557271
DBSOURCE    REFSEQ: accession NM_001126.1
```

RefSeq Protein

```
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI
            staff. The reference sequence was derived from X66503.1.
            Summary: Adenylosuccinate synthetase catalyzes the first
            committed step in the conversion of IMP to AMP.
```

X records: Genome Annotation & Inferred or Predicted
vs
N records: Provisional, Reviewed or Validated

Third Party Annotation (TPA) Database

NCBI FieldGuide

NCBI now accepts the submission of new annotations of existing GenBank sequences.

- Submissions n … d journal.
- Facilitates the … s.

Examples of seq

– Annotation of fe … uences
– Assembled "full

What should not

– Synthetic const … use well-characterized, publicly availab …
– Updates or cha …
– Sequence annotations *without* experimental evidence

1: BK000627 Links
TPA: Homo sapiens equilibrative nucleoside transporter 4 (ENT4) gene, complete cds
gi[25418479]tpg[BK000627.1][25418479]

2: BK000646 Links
TPA: Homo sapiens keratin 5b (K5B) gene, complete cds
gi[25140705]tpg[BK000646.1][25140705]

3: BK000688 Links
TPA: Homo sapiens synovial sarcoma X breakpoint 8 protein (SSX8) mRNA, complete cds
gi[25140703]tpg[BK000688.1][25140703]

4: BK000689 Links
TPA: Homo sapiens synovial sarcoma X breakpoint 9 protein (SSX9) mRNA, complete cds
gi[25140701]tpg[BK000689.1][25140701]

5: BK000686 Links
TPA: Homo sapiens synovial sarcoma X breakpoint 6 protein (SSX6) mRNA, complete cds
gi[25140699]tpg[BK000686.1][25140699]

RefSeq

genomes·transcripts·proteins

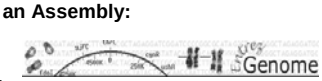
NCBI FieldGuide

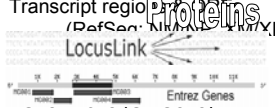
- "Best representative" (reference) sequences
- Standardized nomenclature and record structure
- Added annotation (references, sequence features)

Genomes

Genoming Genome Data on an Assembly:

Genome sequence
(Ref… NW)
Transcript regio…
(RefSeq: NM… XP)





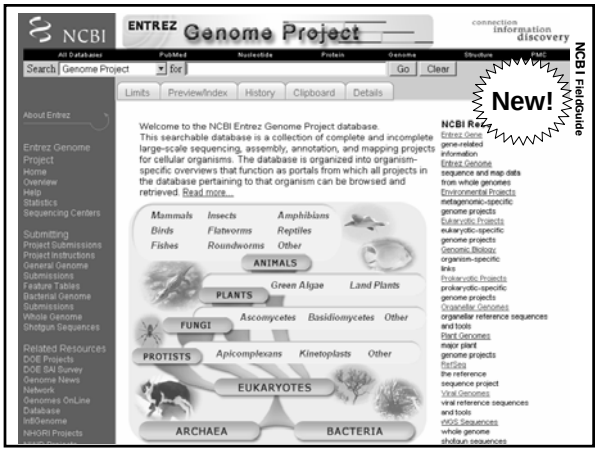
Complete Genomes

as of March 2006

NCBI FieldGuide

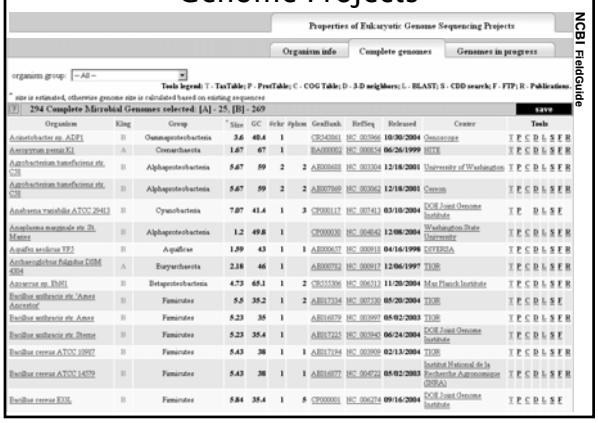
Organelles:
- **Mitochondria (816)**
- **Plastids (50)**
- **Plasmids (871)**
- **Nucleomorphs (3)**

- **Viruses (2286)**
- **Archaebacteria (26)**
- **Eubacteria (285)**
- **Eukaryotes (21complete/86assemblies)**



Genome Projects

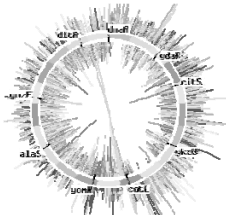
NCBI FieldGuide



Simple Genomes

NCBI FieldGuide

- Full chromosomal sequences are provided
- Genes are annotated
- The annotation can be shown graphically and linked to sequence records



RefSeq Chromosomes: NC_

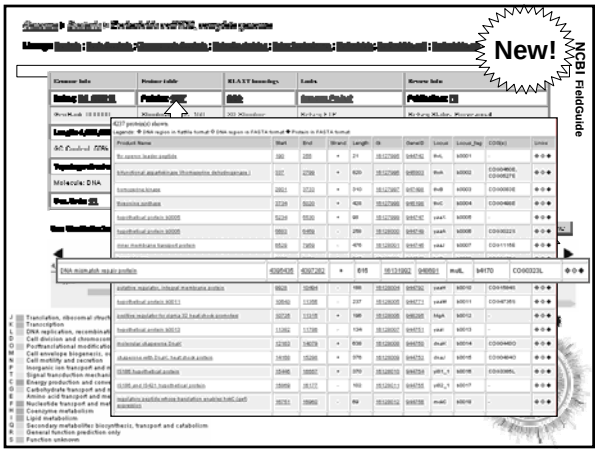
NCBI FieldGuide

```
1:  NC_000913. Reports  Escherichia coli ...[gi:49175990]                    Links
Comments    Features    Sequence

gene     3954631..3956478
         /gene="mutL"
         /locus_tag="b4170"
BASE COUNT    978672 a1011074 c 997153 g 974742 t
ORIGIN
             1 cgtcttcatt gtcagacagc agaatttgta cgcgctgttc ggcttgttgt aatttggcct
            61 gcccctgacg tgccagctgc acgccgcgtt cgaactcgtt cagcgcctct tccagcggca
           121 ggtcgccact ttccagacgg gttacaatct gttccagctc gctcagcgcc ttttcaaagc
           181 tggcgggcgc ctcatttttc ttcggcataa tgaatgtctg actctcaata tttttcgccc
           241 cgtcatggta acggactcag ggcaaatagc aaataacgcg caatggtaag gtgatgtgca
           301 cagcaaagcg atgttagtgg tatacttccg cgcctggatg cagccgcagg tgtgggctgc
           361 tgtatttttc cctatacaag tcgcttaagg cttgccaacg aaccattgcc gccatgaagt
           421 ttatcattaa attgttcccg gaaatcacca tcaaaccca atctgtgcgc ttgcgcttta
           481 taaaaatcct taccgggaac attcgtaacg ttttaaaaca ctatgatgag acgctcgctg
           541 tcgtccgcca ctgggataac atcgaagttc gcgcaaaaaa tgaaaaccag cgtctggcta
           601 ttcgcgacgc tctgacccgt attccgggta tcc                    acgtgc
           661 cgtttaccga catgcacgat attttcgaga aag                    agctgg
           721 aaggcaaaac cttctgcgta cgcgtgaagc gccgtggcaa acatgatttt agctcgattg
           781 atgtggaacg ttacgtcggc ggcggtttaa atcagcatat tgaatccgcg cgcgtgaagc
           841 tgaccaatcc ggatgtgact gtccatctgg aagtggaaga cgatcgtctc ctgctgatta
           901 aaggccgcta cgaaggtatt ggcggtttcc cgatcggcac ccaggaagat gtgctgtcgc
           961 tcatttccgg tggtttcgac tccggtgttt ccagttatat gttgatgcgt cgcggctgcc

        SEHAQWSMAQATTTLADVERTCPQLVKTPPGGTTQSVDLHPATKALKDE"
```

Genome sequence

New!

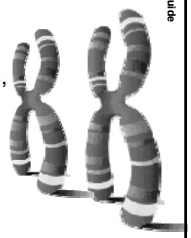
NCBI FieldGuide



Complex Genomes

NCBI FieldGuide

- Sequences are provided complete or we help assemble
- Heavy annotation:
  Genes, transcript regions & ORFs, sequence variations & markers, clones, ESTs, etc.
- The annotation can be shown graphically and linked to other databases using the MapViewer



A database for retrieval and analysis of karyotype data:
Cancer Chromosomes

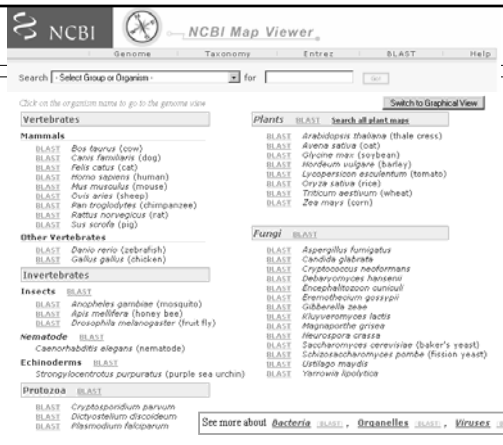
RefSeq Records Contig: NT_ & Chromosome: NC_

NCBI FieldGuide

```
1: NT_003440.. Reports Homo sapiens chro...[gi:51458694]          Links
Comments    Features    Sequence

LOCUS       NT_034400   1065823 bp    DNA     linear   CON  19-AUG-2004
DEFINITION  Homo sapiens chromosome 1 genomic contig.
ACCESSION   NT_034400
VERSION     NT_034400.3  GI:51458694
gene        complement(2548206..2591802)
            /gene="ADSS"
            /db_xref="LocusID:159"
            /db_xref="MIM:103060"
mRNA        complement(join(2548206..2549349,2550998..2551147,
            2555692..2555789,2557339..2557463,2558471..2558625,
            2559881..2560007,2562526..2562607,2563644..2563751,
            2564012..2564078,2572236..2572286,2576516..2576584,
            2577357..2577459,2591326..2591802))
            /gene="ADSS"
CONTIG      join(AL139152.7:1..55543,AL596177.4:1998..91084,
             AL356378.17:1999..202955,AL391904.14:2001..68222,
             AL590667.7:2001..175494,AL359207.7:2001..112707,
             AL365260.11:2001..114412,complement(AL445591.10:1..138092),
             BX537254.7:2001..121309)
//
            /db_xref="MIM:103060"
```

Annotation of Gene, mRNA, CDS, and other features

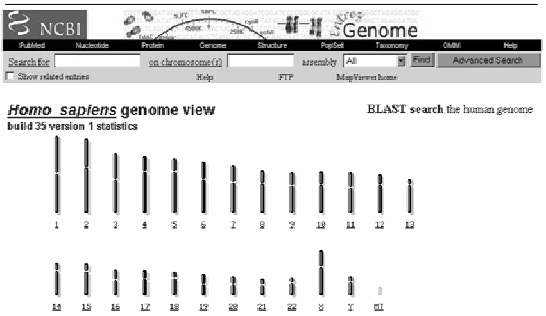
Ordering of draft sequences



NCBI FieldGuide

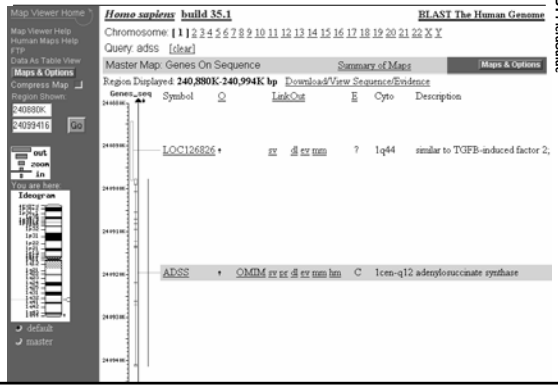
Higher Genome MapViews

NCBI FieldGuide



Higher Genome MapViews

NCBI FieldGuide



Gene Expression Databases

NCBI FieldGuide

New!

A new database for localization of proteins in Mouse Brains:
GenSat
A new database for information on Expression Reagents:
Probe

NCBI FieldGuide

Submit and update data

Query the database:
- **gene identifiers**
- **field information**
- **sequence**

Browse datasets

Download data

GEO Gene Expression Omnibus

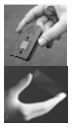
NCBI FieldGuide

Submitted by Manufacturer*

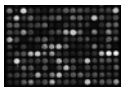
Submitted by Experimentalists

Curated by NCBI

GPL
Platform descriptions



GSM
Raw/processed spot intensities from a single slide/chip



GSE
Grouping of slide/chip data
"a single experiment"

GDS
Grouping of experiments

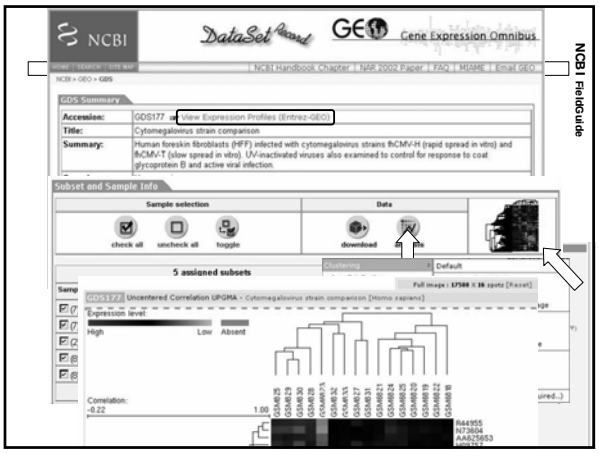
Entrez GEO

Entrez GEO Datasets

Entrez GEO DataSets

NCBI FieldGuide

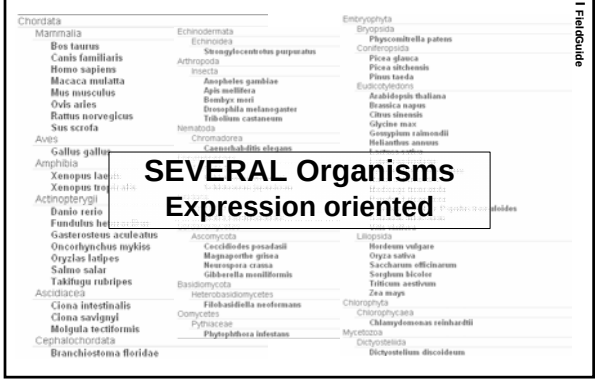
GDS177: CMV infection of HFF cells



NCBI FieldGuide

UniGene Collections

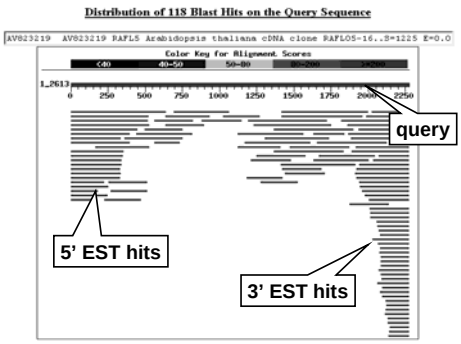
as of January 2006



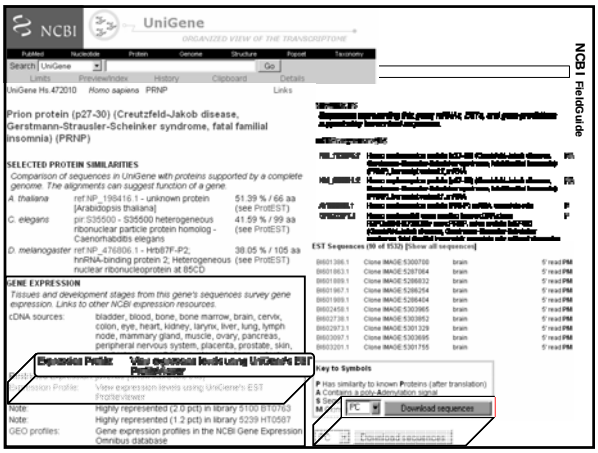
SEVERAL Organisms
Expression oriented

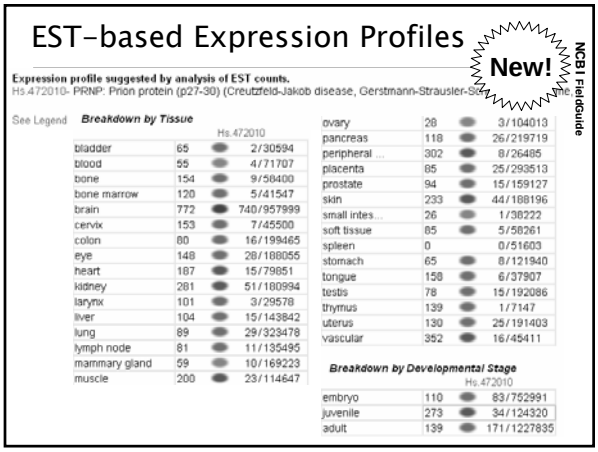
NCBI FieldGuide

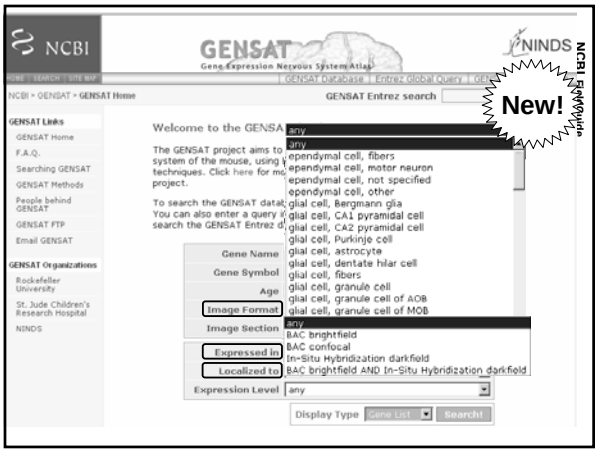
A Cluster of ESTs:
Arabidopsis serine protease

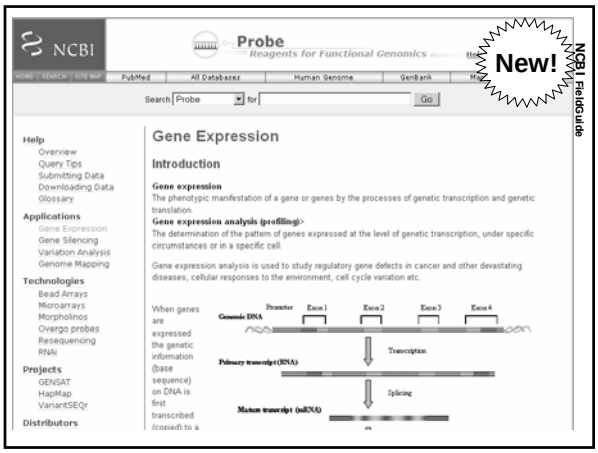


NCBI FieldGuide


NCBI
UniGene
Prion protein (p27-30) (Creutzfeld-Jakob disease, Gerstmann-Strausler-Scheinker syndrome, fatal familial insomnia) (PRNP)
SELECTED PROTEIN SIMILARITIES
GENE EXPRESSION
Download sequences
NCBI FieldGuide


EST-based Expression Profiles
New!
Expression profile suggested by analysis of EST counts.
See Legend
Breakdown by Tissue
Breakdown by Developmental Stage
NCBI FieldGuide


NCBI
GENSAT
Gene Expression Nervous System Atlas
NINDS
New!
GENSAT Entrez search
GENSAT Links
Welcome to the GENSA
Gene Name
Gene Symbol
Age
Image Format
Image Section
Expressed in
Localized to
Expression Level
Display Type
Search!
NCBI FieldGuide

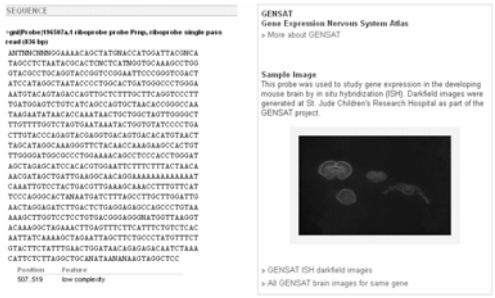


NCBI FieldGuide

Probe: Expression probes

Pr196507.1 Links

Ribonucleic acid probe (riboprobe) Prnp for Mus musculus gene prion protein (Prnp). Has been used in the GENSAT project for in situ hybridization.

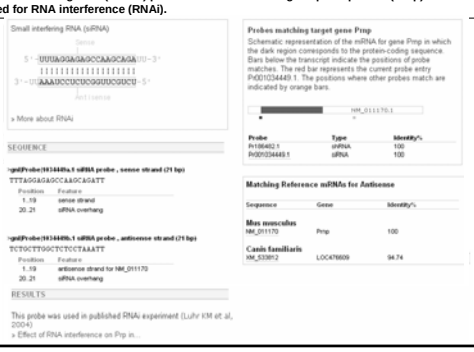


NCBI FieldGuide

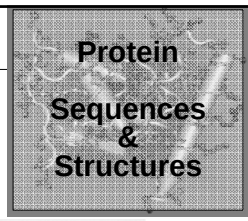
Probe: siRNAs & shRNAs

Pr001034449.1 Links

Small interfering RNA (siRNA) probe for Mus musculus gene prion protein (Prnp). Has been used for RNA interference (RNAi).



NCBI FieldGuide


Protein
Sequences
&
Structures

NCBI FieldGuide

Linking Protein Sequence, Structure and Function

NCBI FieldGuide

Protein — Protein sequences

Conserved Domain — **CDD:** Conserved functional domains in proteins represented by a PSSM

RPS-BLAST, CDART

Structure — **MMDB:** Experimentally-derived 3D structure records from PDB

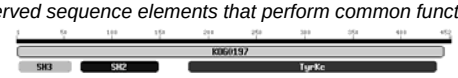
3D Domain — Compact structural domains of protein folds

Sequence-based Neighbors: Domain Neighbors

NCBI FieldGuide

-Conserved Domains-

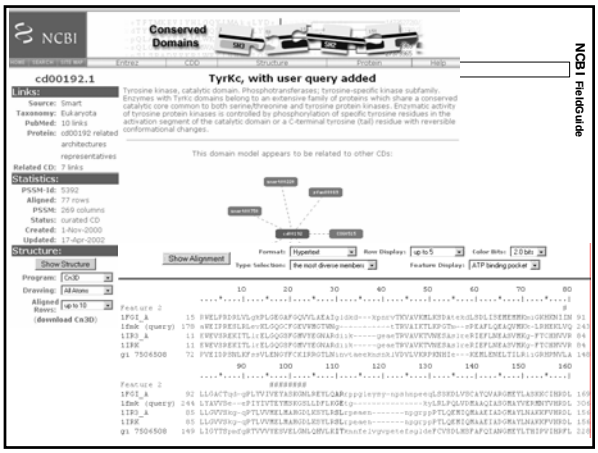
conserved sequence elements that perform common functions



Curation of protein multiple sequence alignments with known similar function by conversion to Position-Specific Scoring Matrices

```
                    10        20        30        40        50        60
                ....*....|....*....|....*....|....*....|....*....|....*....|
consensus   1 KWEIPREDLTLGKKLGEGAFGEVYKGTLKGkgd---nkSIDVAVKTLKEDASEeqIKEFL  57
1FGI_A      1 aWEIPRESLRLEVKLQQGCFGEVWMGTWNG--------TTRVAIKTLKPGTMS--PEAFL  311
1BYG_A      1 RWELPRDRLVLgkPLGEGAFGQVYLAEAIglgkdkpnrvTKVAVKMLKSDAtedkLSLDI  74
gi 125135   1 GWALNMKELKLlqTIGKGEFGDVMLGDYRg---------NKVAVKCIKNDAt---AQAFL  62
gi 125702   1 KYEIPRTDLTLkhKLGGGQYGEVYEGVWKky--------sLTVAVKTLKEDTm--eVEEFL  284
gi 1174437  1 KWEIPRSELTIlrKLGRGNFGEVFYGKWRn--------sIDVAVKTLREGTm--sTAAFL  325
```

"Reverse-Position Specific" Sequence Comparisons (RPS-BLAST)
a.k.a.
"Conserved Domain Database" (CDD) Search

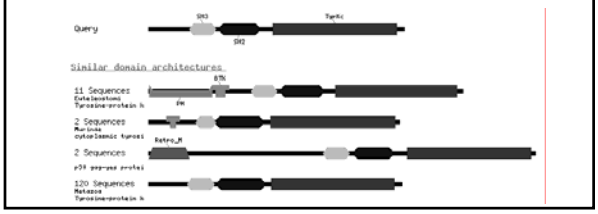


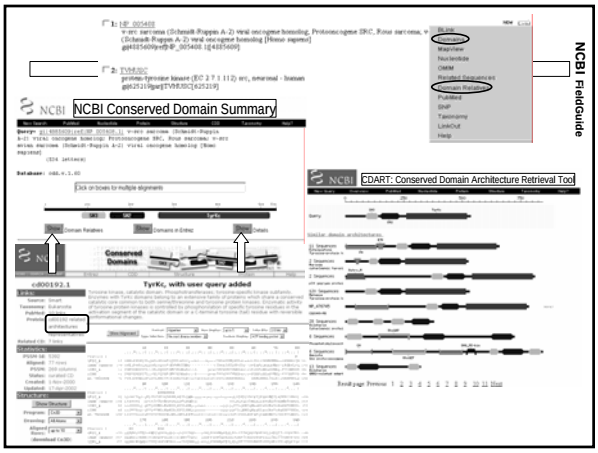
Sequence-based Neighbors: Domain Relatives

"Conserved Domain Architecture Retrieval Tool" (CDART)

Modular Architecture of Domains

- Cartoon descriptions of protein domain organization on the primary sequence
- Allows for comparison with other proteins with the same Domain





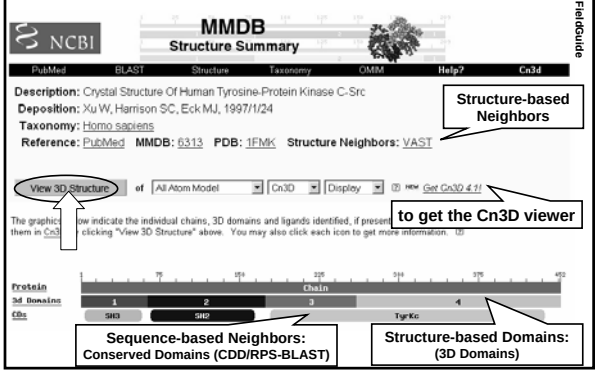
Entrez Structure: Molecular Modeling Database

NCBI FieldGuide

- Derived from experimentally determined PDB records
- Data is added to PDB records including:
  - Addition of explicit chemical bonding information
  - Validation and indexing of sequence
  - Inclusion of Taxonomy, Citation, and other information
  - Conversion to ASN.1 data description language
- Searching the Structure Databases:
  - Keyword search by Entrez
  - Sequence search by BLAST or BLink
  - Domain search by CDD/RPS-BLAST
  - Structure search by VAST

Structure Summary Page

NCBI FieldGuide



Entrez PubChem

New!

NCBI FieldGuide

PC Compound

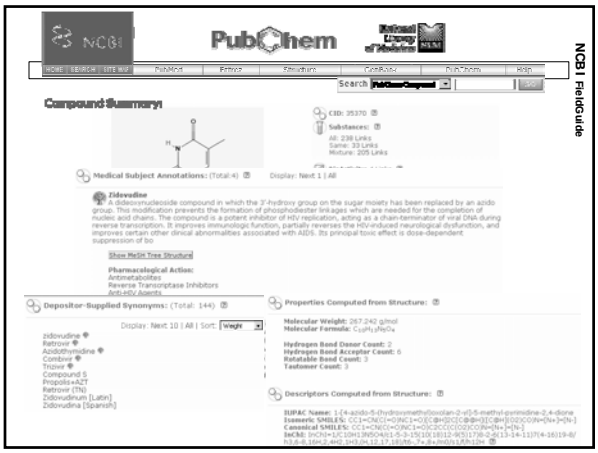
Derived database of known chemicals from PC Substance records

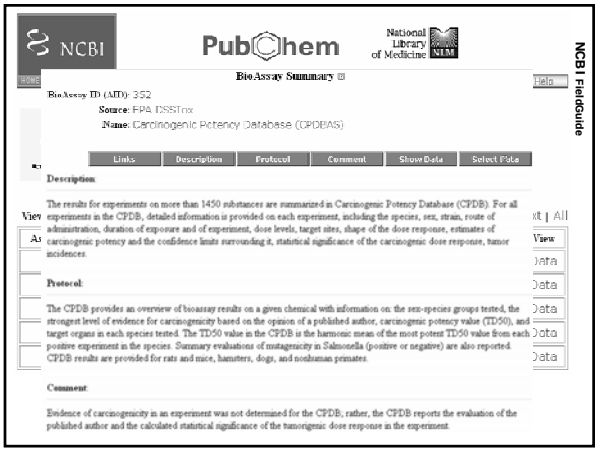
PC Substance

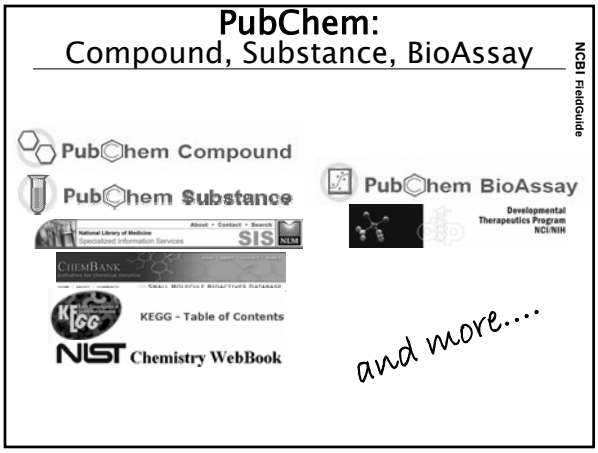
Primary database of chemical samples

PC BioAssay

Primary database of bioactivity screens of samples in PC Substance


NCBI
PubChem
National Library of Medicine
NLM
Compound Summary
Medical Subject Annotations: (Total:4)
Zidovudine
Show MeSH Tree Structure
Pharmacological Action:
Antimetabolites
Reverse Transcriptase Inhibitors
Anti-HIV Agents
Depositor-Supplied Synonyms: (Total: 144)
Properties Computed from Structure:
Descriptors Computed from Structure:
NCBI FieldGuide


NCBI
PubChem
National Library of Medicine
NLM
BioAssay Summary
BioAssay ID (AID): 352
Source: EPA CSSTox
Name: Carcinogenic Potency Database (CPDBAS)
Links
Description
Protocol
Comment
Show Data
Select Data
Description:
The results for experiments on more than 1450 substances are summarized in Carcinogenic Potency Database (CPDB). For all experiments in the CPDB, detailed information is provided on each experiment, including the species, sex, strain, route of administration, duration of exposure and of experiment, dose levels, target sites, shape of the dose response, estimates of carcinogenic potency and the confidence limits surrounding it, statistical significance of the carcinogenic dose response, tumor incidences.
Protocol:
The CPDB provides an overview of bioassay results on a given chemical with information on: the sex-species groups tested, the strongest level of evidence for carcinogenicity based on the opinion of a published author, carcinogenic potency value (TD50), and target organs in each species tested. The TD50 value in the CPDB is the harmonic mean of the most potent TD50 value from each positive experiment in the species. Summary evaluations of mutagenicity in Salmonella (positive or negative) are also reported. CPDB results are provided for rats and mice, hamsters, dogs, and nonhuman primates.
Comment:
Evidence of carcinogenicity in an experiment was not determined for the CPDB; rather, the CPDB reports the evaluation of the published author and the calculated statistical significance of the tumorigenic dose response in the experiment.
NCBI FieldGuide


PubChem:
Compound, Substance, BioAssay
NCBI FieldGuide
PubChem Compound
PubChem Substance
PubChem BioAssay
Developmental Therapeutics Program NCI/NIH
SIS
ChemBank
KEGG - Table of Contents
NIST Chemistry WebBook
and more....

Entrez Gene

The Gene Summary Database

Summary pages
of curated information about genetic loci
for organisms in the RefSeq project.

- ►Graphics
- ►Gene information
- ►Bibliography (PubMed links)
- ►General gene information
- ►NCBI Reference Sequences
- ►Related sequences
- ►Additional Links

NCBI FieldGuide

Entrez Genes

glucose-6-phosphate dehydrogenase

H.sapiens & B.taurus
G6PD

R.norvegicus
G6pdx

M.musculus
G6pd1, G6pdx

D.melanogaster
Zw

A.thalia
At5g35970

S.pombe
SPAC3C7.13c
SPAC9.01
SPCC794.01c

B.anthracis
BA_3932

H.pylori
HP1101

zwf
E.coli, Salmonella, Shigella, Yersinia, Neisseria....

NCBI FieldGuide

Default Display

GENE SYMBOL & name [*Organism*]

- **►Gene information:**
  Gene type
  Gene name
  Gene description
  RefSeq status
  Organism
  Lineage
  Gene Aliases
  Summary
  General protein information
- **►Graphics:**
  Transcripts and products
  Genomic context
- **►Bibliography (PubMed links)**
  GeneRifs
- **►General gene information:**
  Gene Ontology
  Homology (Mouse, Rat, Human)
  Phenotypes
  Sequence Tagged Sites
  Pathways
- **►NCBI Reference Sequences**
  mRNA sequence
  Source sequence
  Product
  Conserved Domains
- **►Related sequences**
  (genomic, mRNA & protein)
- **►Additional Links**

G6PD glucose-6-phosphate dehydrogenase [Homo

1: G6PD glucose-6-phosphate dehydrogenase [*Homo sapien*

GeneID: 2539 Primary source: HGNC:4057

Summary

Genomic regions, transcripts, and products

Bibliography Gene References into Function (GeneRIF): Submit

General gene information

General protein information

Related Sequences

	Nucleotide	Protein
Genomic:	AF277315	AAL27011
Genomic:	AY158096	AAN76363
Genomic:	AY158097	AAN76368
Genomic:	AY158098	AAN76369
Genomic:	X53815	CAA37811
Genomic:	X55445	CAA39089
mRNA	BC000337	AAH00337
mRNA	M19866	AAA51501
mRNA	M21248	AAA52509
mRNA	M27940	AAA52504
mRNA	S58359	AAB26169
mRNA	X03674	CAA27309
	None	P11413

Additional Links

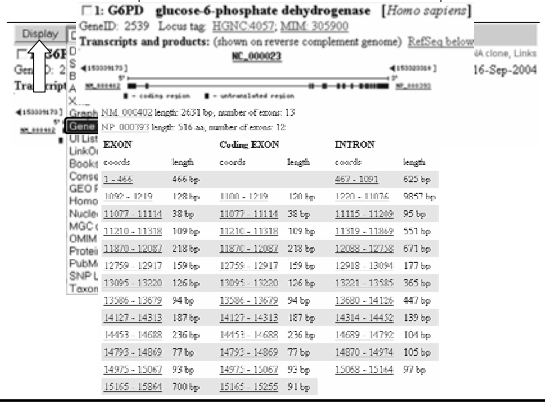
UniGene Hs.461047
MIM 305900
GeneTests for MIM: 305900
PharmGKB PA28469
HPRD 02377
Favism Research Papers Favism Research Papers
Integrated X Chromosome Database (IXDB) Integrated X Chromosome Database (IXDB)

Table Of Contents
Summary
Genomic regions, transcripts...
Genomic context
Bibliography
HIV-1 protein interactions
Interactions
General gene information
General protein information
Reference Sequences
Related Sequences
Additional Links

Links
MGC cDNA clone
Conserved Domains
Genome
GEO Profiles
HomoloGene
Map Viewer
Nucleotide
OMIM
Full text in PMC
Probe
Protein
PubMed
PubMed (GeneRIF)
SNP
SNP: Genotype
SNP: GeneView
Taxonomy
UniSTS
AceView
CCDS
Ensembl
Evidence Viewer
Favism Research Papers
GDB
GeneTests for MIM: 305900
HGMD
HGMC
HPRD
Integrated X Chromosome...
KEGG
MGC
ModelMaker
PharmGKB
Reactome
UCSC
UniGene
LinkOut

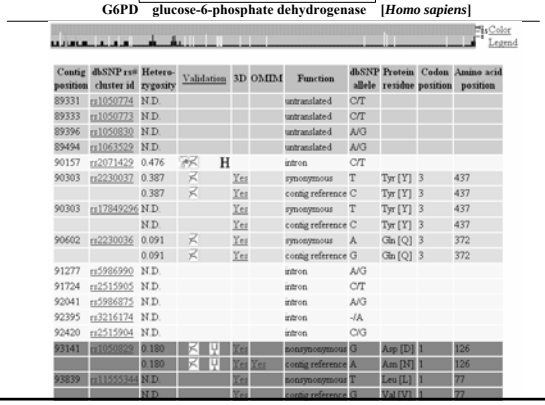
Entrez Gene Info

Gene Table Display



NCBI FieldGuide

SNP: GeneView



Contig position	dbSNP rs#	Hetero-zygosity	Validation	3D	OMIM	Function	dbSNP allele	Protein residue	Codon position	Amino acid position
89331	rs1050774	N.D.				untranslated	C/T			
89333	rs1050773	N.D.				untranslated	C/T			
89396	rs1050830	N.D.				untranslated	A/G			
89494	rs1063529	N.D.				untranslated	A/G			
90157	rs2071429	0.476		H		intron	C/T			
90303	rs2230037	0.387			Yes	synonymous	T	Tyr [Y]	3	437
		0.387			Yes	contig reference	C	Tyr [Y]	3	437
90303	rs17849296	N.D.			Yes	synonymous	T	Tyr [Y]	3	437
		N.D.			Yes	contig reference	C	Tyr [Y]	3	437
90602	rs2230036	0.091			Yes	synonymous	A	Gln [Q]	3	372
		0.091			Yes	contig reference	G	Gln [Q]	3	372
91277	rs5986990	N.D.				intron	A/G			
91724	rs2515905	N.D.				intron	C/T			
92041	rs5986875	N.D.				intron	A/G			
92395	rs3216174	N.D.				intron	-/A			
92420	rs2515904	N.D.				intron	C/G			
93141	rs1050829	0.180			Yes	nonsynonymous	G	Asp [D]	1	126
		0.180			Yes Yes	contig reference	A	Asn [N]	1	126
93839	rs11555344	N.D.			Yes	nonsynonymous	T	Leu [L]	1	77
		N.D.			Yes	contig reference	G	Val [V]	1	77

NCBI FieldGuide

FTP Downloads

NCBI FieldGuide

Help for Programmers

NCBI | FieldGuide

NCBI Toolbox: In-house source code useful for incorporating NCBI-like functionality into their programs.
Three main parts: Data Model, Data Encoding and Programming Libraries.

- Examples: BLAST, Cn3D, Sequin, Data format conversion scripts

http://www.ncbi.nlm.nih.gov/IEB/ToolBox/index.cgi

E-Utilities: Guidelines for Entrez "URL calls" used to access data.
Designed for use in scripts.

- Examples: ESearch, EPost, ESummary, EFetch and ELink

http://www.ncbi.nih.gov/entrez/query/static/eutils_help.html

Caution: *Overuse may result in blocked IPs!*

Intermission

NCBI | FieldGuide