常用生物医学数据库与分析软件介绍

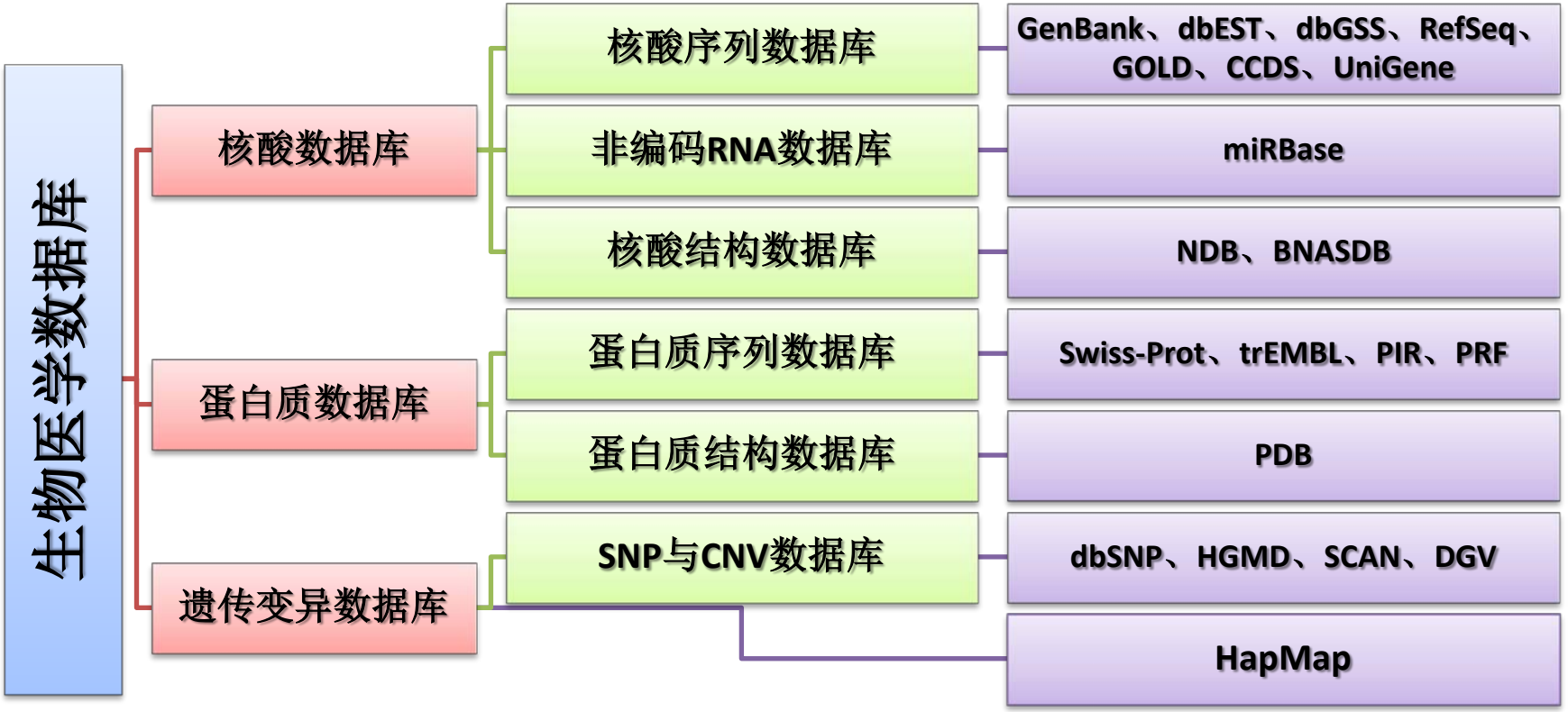
宫滨生

生物信息科学与技术学院

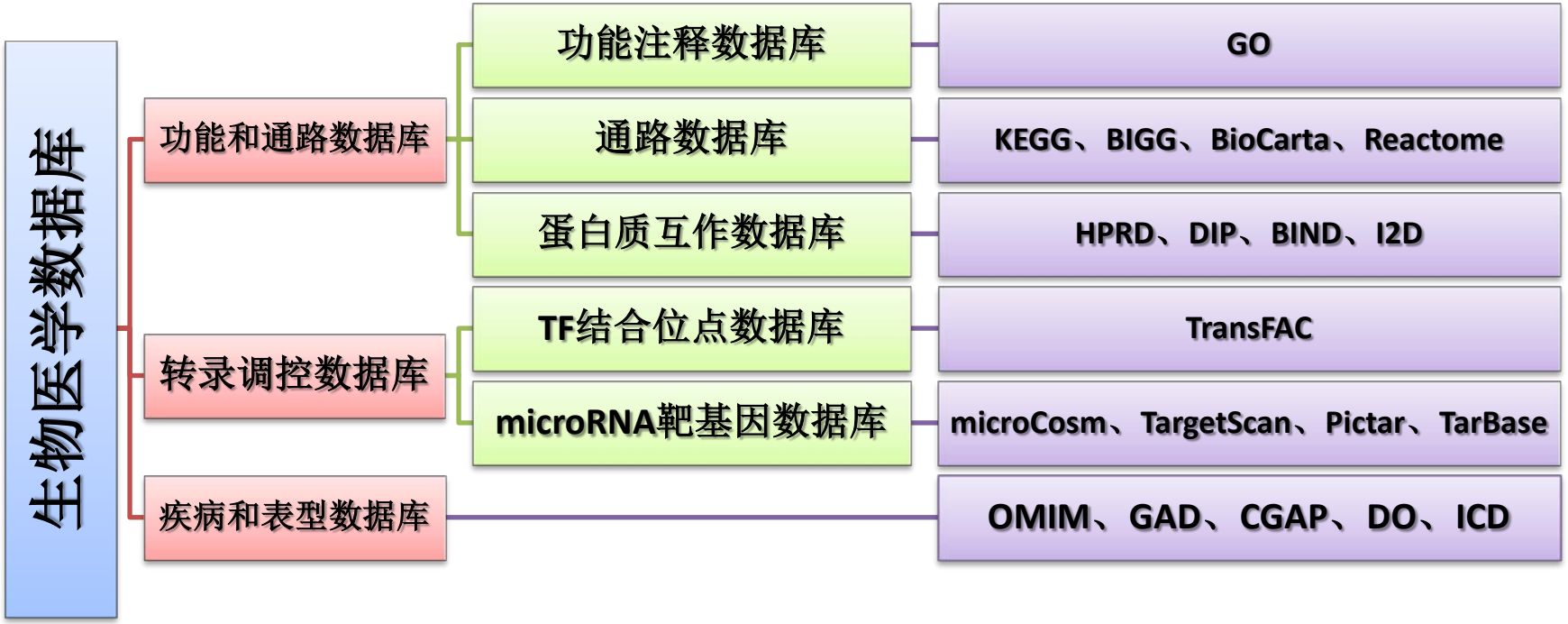
生物医学数据库概览


生物医学数据库
核酸数据库
核酸序列数据库
GenBank、dbEST、dbGSS、RefSeq、GOLD、CCDS、UniGene
非编码RNA数据库
miRBase
核酸结构数据库
NDB、BNASDB
蛋白质数据库
蛋白质序列数据库
Swiss-Prot、trEMBL、PIR、PRF
蛋白质结构数据库
PDB
表达谱数据库
基因表达谱数据库
GEO、SMD、ArrayExpress
遗传变异数据库
SNP与CNV数据库
dbSNP、HGMD、SCAN、DGV
HapMap
功能和通路数据库
功能注释数据库
GO
通路数据库
KEGG、BIGG、BioCarta、Reactome
蛋白质互作数据库
HPRD、DIP、BIND、I2D
转录调控数据库
TF结合位点数据库
TransFAC
microRNA靶基因数据库
microCosm、TargetScan、Pictar、TarBase
疾病和表型数据库
OMIM、GAD、CGAP、DO、ICD
新一代测序数据库
DNA-seq、RNA-seq、ChIP-seq
SRA、GEO、hmChIP
整合数据库
Ensembl
UCSC
其他数据库
文献数据库
PubMed、MeSH
实验相关数据库
dbGaP
同源数据库
COG/KOG、CDD
小分子数据库
DrugBank、cMAP、PubChem

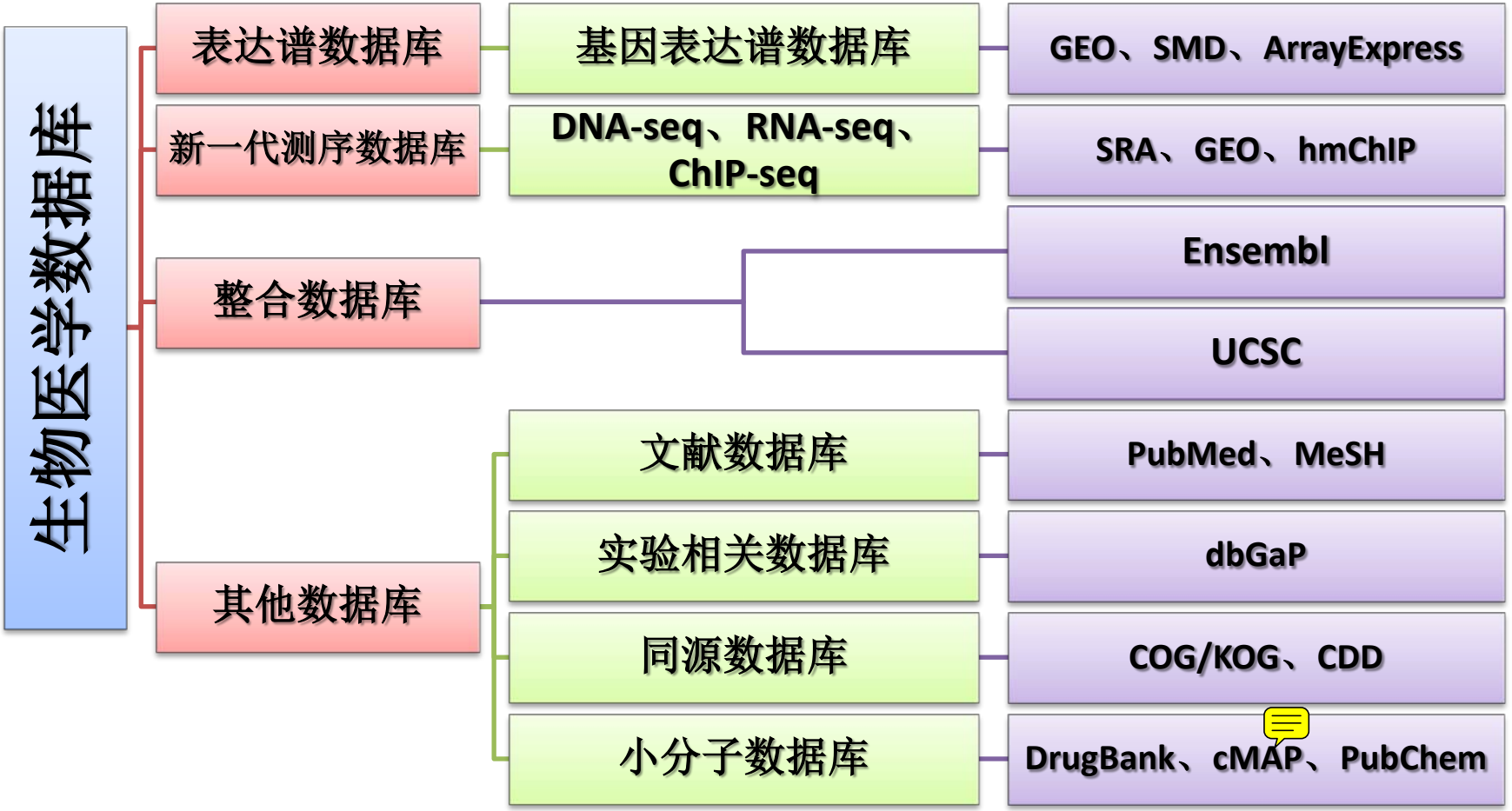
生物医学数据库概览


生物医学数据库
核酸数据库
蛋白质数据库
遗传变异数据库
核酸序列数据库
非编码RNA数据库
核酸结构数据库
蛋白质序列数据库
蛋白质结构数据库
SNP与CNV数据库
GenBank、dbEST、dbGSS、RefSeq、GOLD、CCDS、UniGene
miRBase
NDB、BNASDB
Swiss-Prot、trEMBL、PIR、PRF
PDB
dbSNP、HGMD、SCAN、DGV
HapMap

生物医学数据库概览


生物医学数据库
功能和通路数据库
转录调控数据库
疾病和表型数据库
功能注释数据库
通路数据库
蛋白质互作数据库
TF结合位点数据库
microRNA靶基因数据库
GO
KEGG、BIGG、BioCarta、Reactome
HPRD、DIP、BIND、I2D
TransFAC
microCosm、TargetScan、Pictar、TarBase
OMIM、GAD、CGAP、DO、ICD

生物医学数据库概览


生物医学数据库
表达谱数据库
基因表达谱数据库
GEO、SMD、ArrayExpress
新一代测序数据库
DNA-seq、RNA-seq、ChIP-seq
SRA、GEO、hmChIP
整合数据库
Ensembl
UCSC
其他数据库
文献数据库
PubMed、MeSH
实验相关数据库
dbGaP
同源数据库
COG/KOG、CDD
小分子数据库
DrugBank、cMAP、PubChem

NCBI National Center for Biotechnology Information

Search All Databases

Search Clear

National Center for Biotechnology Information

- NCBI Home
- Site Map (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

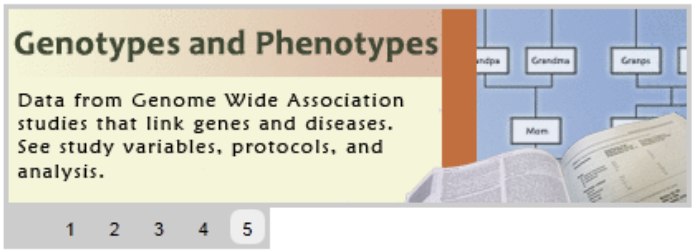
Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | RSS Feeds

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-To's: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases



Popular Resources

- BLAST
- Bookshelf
- Gene
- Genome
- Nucleotide
- OMIM
- Protein
- PubChem
- PubMed
- PubMed Central
- SNP

NCBI News

New NCBI News Issue

06 Jul 2011

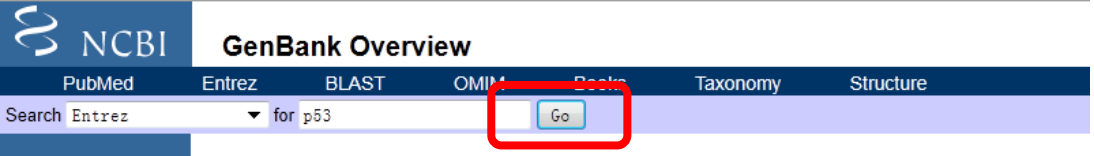
Information on the redesigned PopSet resource, as well as new

Preliminary genomic assemblies from two isolates from the European E. coli outbreak now available

07 Jun 2011

Preliminary genomic assemblies of two isolates are in the Nucleotide

More...



NCBI Home

NCBI Site Map

GenBank Submissions Handbook

Submit to GenBank

Submit an update

Search GenBank

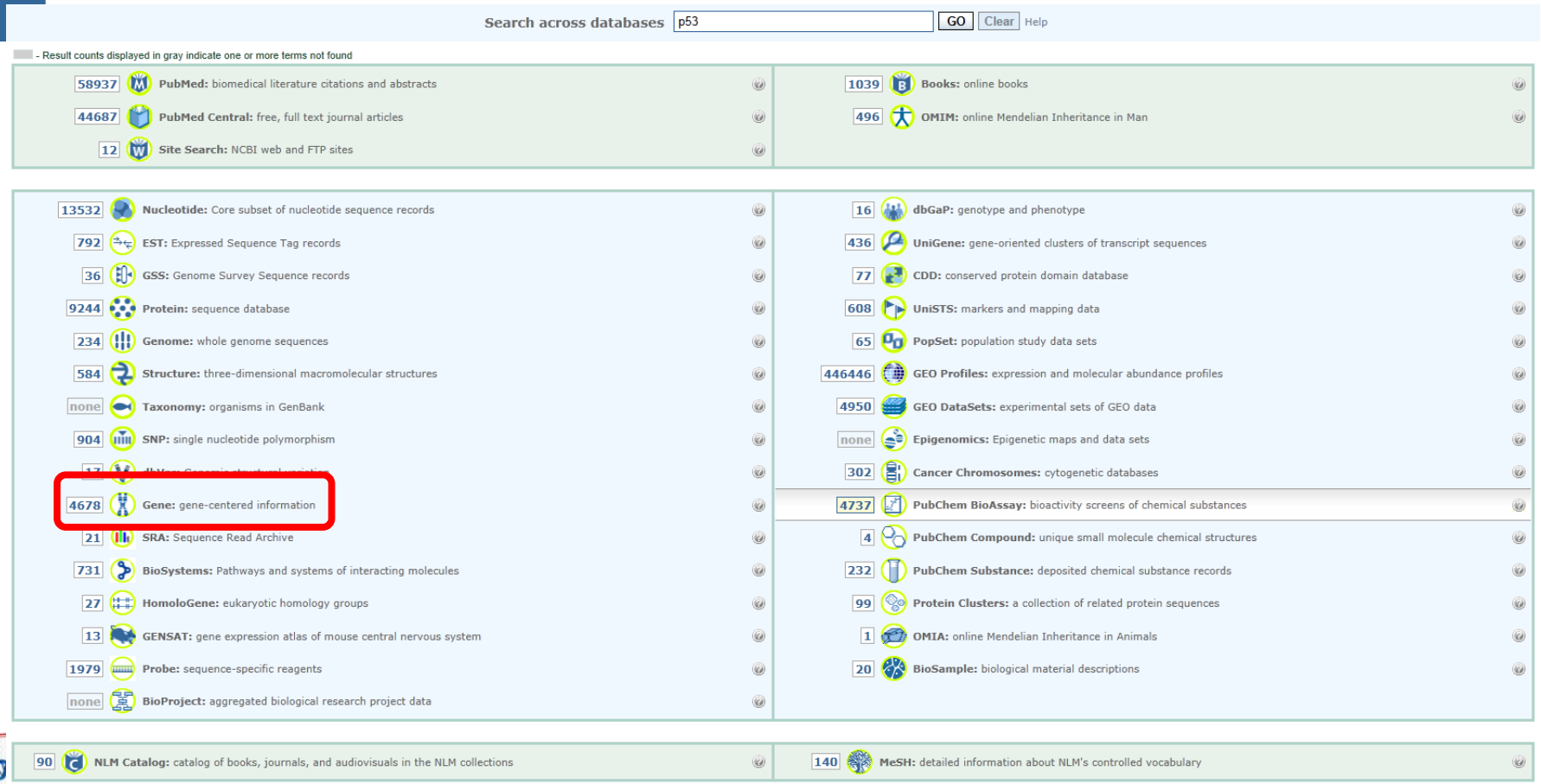
GenBank and RefSeq: a comparison

BLAST

What is GenBank?

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*. 2011 Jan;39(Database issue):D32-7). There are approximately 126,551,501,141 bases in 135,440,924 sequence records in the traditional GenBank divisions and 191,401,393,188 bases in 62,715,288 sequence records in the WGS division as of April 2011.

The complete release notes for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.



NCBI Resources How To

Gene

Genes and mapped phenotypes

Search: Gene | Save search | Limits | Advanced search | Help

p53 | Search | Clear

Display Settings: Summary, 20 per page, Sorted by Relevance

Results: 1 to 20 of 4678

1. **p53**
Official Symbol: p53 and **Name:** CG33336 gene product from transcript CG33336-RB [*Drosophila melanogaster*]
Other Aliases: Dmel_CG33336, CG10873, CG31325, CG33336, D-**p53**, Dm-**P53**, DmP53, Dmel\CG33336, Dmp53, Dp53, dmp53, dp53, prac
Other Designations: CG33336-PA; CG33336-PB; CG33336-PC; Dmp53; **p53**-PA; **p53**-PB; **p53**-PC; **p53**-like regulator of apoptosis and cell cycle
Chromosome: 3R; **Location:** 94D10-94D10
Annotation: Chromosome 3R, NT_033777.2 (18875379..18879804, complement)
ID: 2768677

2. **p53**
p53 gene product [*Drosophila melanogaster*]
Other Aliases: CG10873, Dmp53, dp53
Other Designations: CG10873-PA
Chromosome: 3R; **Location:** 94D10-94D10
Annotation: Chromosome 3R, NT_033777 (18866029..18869867, complement)
This record was discontinued
ID: 42722

3. **p53**
p53 tumor suppressor homolog [*Bombyx mori*]
Other Designations: **p53** tumor suppressor homologue
ID: 100384887

4. **P53**
hypothetical protein [*Bacteriophage APSE-2*]
Other Aliases: APSE242
Annotation: NC_011551.1 (38386..39303, complement)
ID: 7020953

5. **P53**
APSE-2 prophage; hypothetical [*Candidatus Hamiltonella defensa 5AT (Acyrthosiphon pisum)*]
Other Aliases: HDEF_1662
Genomic context: Chromosome
Annotation: NC_012751.1 (1511504..1512421, complement)
ID: 7951260

6. TP53
Official Symbol: TP53 and **Name:** tumor protein **p53** [*Homo sapiens*]
Other Aliases: FLJ92943, LFS1, **P53**, TRP53
Other Designations: OTTHUMP00000221333; OTTHUMP00000221334; OTTHUMP00000221336; OTTHUMP00000221337; OTTHUMP00000221
Chromosome: 17; **Location:** 17p13.1

Bioinformatics

期生物信息学培训班

Display Settings: Full Report Send to:

TP53 tumor protein p53 [*Homo sapiens*]

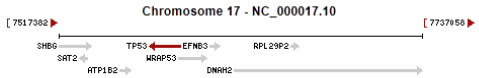
Gene ID: 7157, updated on 17-Jul-2011

Summary

Official Symbol TP53 provided by HGNC
Official Full Name tumor protein p53 provided by HGNC
Primary source HGNC:11998
See related Ensembl:ENSG00000141510; HPRD:01859; MIM:191170
Gene type protein coding
RefSeq status REVIEWED
Organism Homo sapiens
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as P53; LFS1; TRP53; FLJ92943
Summary This gene encodes tumor protein p53, which responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. p53 protein is expressed at low level in normal cells and at a high level in a variety of transformed cell lines, where it's believed to contribute to transformation and malignancy. p53 is a DNA-binding protein containing transcription activation, DNA-binding, and oligomerization domains. It is postulated to bind to a p53-binding site and activate expression of downstream genes that inhibit growth and/or invasion, and thus function as a tumor suppressor. Mutants of p53 that frequently occur in a number of different human cancers fail to bind the consensus DNA binding site, and hence cause the loss of tumor suppressor activity. Alterations of this gene occur not only as somatic mutations in human malignancies, but also as germline mutations in some cancer-prone families with Li-Fraumeni syndrome. Multiple p53 variants due to alternative promoters and multiple alternative splicing have been found. These variants encode distinct isoforms, which can regulate p53 transcriptional activity. [provided by RefSeq]

Genomic context

chromosome: 17; **Location:** 17p13.1 See TP53 in MapViewer



Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence NC_000017 chromosome

Go to nucleotide Graphics FASTA GenBank

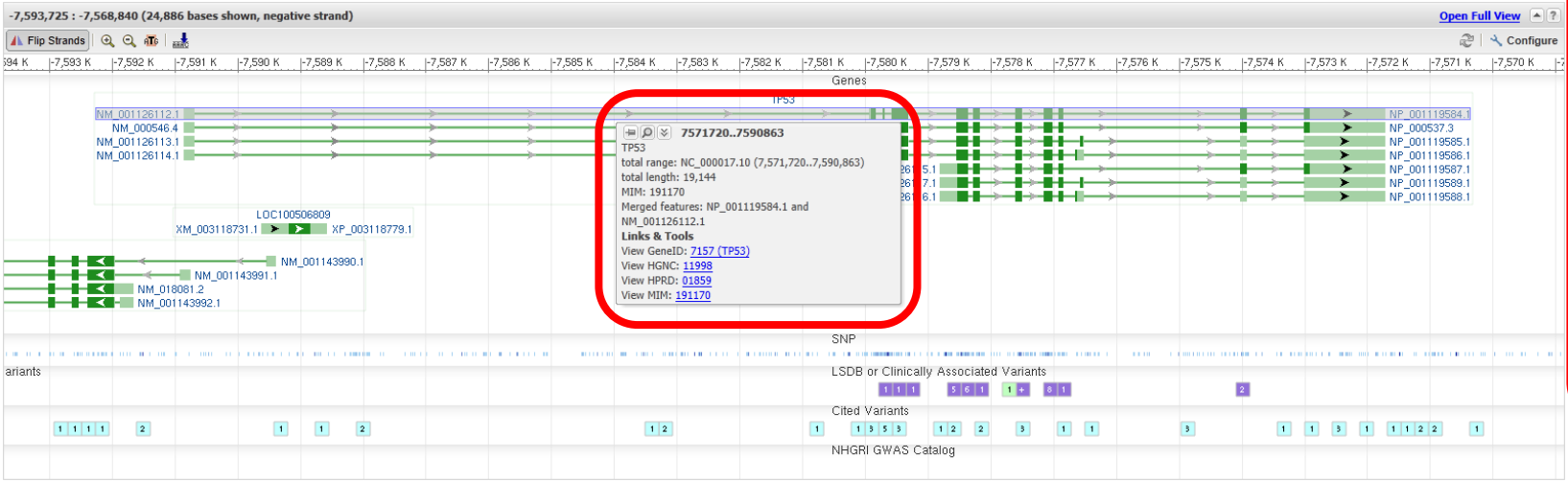


Table of contents

Links

生物信息学培训班

Table of contents

Links

Order cDNA clone
BioAssay
BioAssay, by Gene target
BioProjects
BioSystems
Books
CCDS
Conserved Domains
dbVar
EST
Full text in PMC
GEO Profiles
Genome
HomoloGene
Map Viewer
Nucleotide
OMIM
Peptidome
Probe
Protein
PubChem Compound
PubChem Substance
PubMed
PubMed (GeneRIF)
PubMed (OMIM)
RefSeq Proteins
RefSeq RNAs
RefSeqGene
SNP
SNP: GeneView
SNP: Genotype
SNP: VarView
Taxonomy
UniGene
UniSTS

Links to other resources

AceView
Colon.html
Ensembl
Evidence Viewer
GeneTests for MIM: 191170
HGNC
HPRD
HuGE Navigator
KEGG
MGC
ModelMaker
PharmGKB
Reactome
UCSC
p53 Mutation Database
p53.html
www-p53.iarc.fr/
www.lf2.cuni.cz/projects/germline_mut_p53.htm
www.umd.be:2072/

General information

About Gene
FAQ
FTP site
Help
My NCBI help
NCBI Handbook
Statistics

Related sites

BLAST
Entrez Genome
Genome Project
Genomic Biology
GEO
HomoloGene
Map Viewer
OMIM
Probe
RefSeq
UniGene
UniSTS

Feedback

Contact Help Desk
Submit Correction
Submit GeneRIF

Subscription

RefSeq
Gene
Map Viewer

Recent activity

Turn Off Clear

TP53 tumor protein p53 [Homo sapiens] Gene

p53 (4678) Gene

Homo sapiens origin recognition complex, subunit 2 (ORC2), transcript variant Nucleotide

p53 (13532) Nucleotide

p53 (58937) PubMed

NCBI
National Center for Biotechnology Information

Search All Databases
Search Clear

NCBI Home
Site Map (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

DNA & RNA

All | Databases | Downloads | Submissions | Tools | How To

DATABASES

BioProject (formerly Genome Project)

A collection of genomics, functional genomics, and genetics studies and links to their resulting datasets. This resource describes project scope, material, and objectives and provides a mechanism to retrieve datasets that are often difficult to find due to inconsistent annotation, multiple independent submissions, and the varied nature of diverse data types which are often stored in different databases.

BioSample

The BioSample database contains descriptions of biological source materials used in experimental assays.

Consensus CDS (CCDS)

A collaborative effort to identify a core set of human and mouse protein coding regions that are consistently annotated and of high quality.

Database of Expressed Sequence Tags (dbEST)

Quick Links

BLAST (Stand-alone)
Basic Local Alignment Search Tool (BLAST)
BioProject (formerly Genome Project)
E-Utilities
GenBank
GenBank: BankIt
GenBank: Sequin
GenBank: tbl2asn
Genome Workbench
Nucleotide Database
PopSet
Primer-BLAST
ProSplign
RefSeqGene
Reference Sequence (RefSeq)
Sequence Read Archive (SRA)

DNA & RNA

All | **Databases** | Downloads | Submissions | Tools | How To

DATABASES

BioProject (formerly Genome Project)

A collection of genomics, functional genomics, and genetics studies and links to their resulting datasets. This resource describes project scope, material, and objectives and provides a mechanism to retrieve datasets that are often difficult to find due to inconsistent annotation, multiple independent submissions, and the varied nature of diverse data types which are often stored in different databases.

BioSample

The BioSample database contains descriptions of biological source materials used in experimental assays.

Consensus CDS (CCDS)

A collaborative effort to identify a core set of human and mouse protein coding regions that are consistently annotated and of high quality.

Database of Expressed Sequence Tags (dbEST)

A divison of GenBank that contains short single-pass reads of cDNA (transcript) sequences. dbEST can be searched directly through the Nucleotide EST Database.

Database of Genome Survey Sequences (dbGSS)

A division of GenBank that contains short single-pass reads of genomic DNA. dbGSS can be searched directly through the Nucleotide GSS Database.

GenBank

The NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis. GenBank consists of several divisions, most of which can be accessed through the Nucleotide database. The exceptions are the EST and GSS divisions, which are accessed through the Nucleotide EST and Nucleotide GSS databases, respectively.

All | Databases | Downloads | Submissions | **Tools** | How To

TOOLS

Basic Local Alignment Search Tool (BLAST)

Finds regions of local similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as to help identify members of gene families.

Batch Entrez

Allows you to retrieve records from many Entrez databases by uploading a file of GI or accession numbers from the Nucleotide or Protein databases, or a file of unique identifiers from other Entrez databases. Search results can be saved in various formats directly to a local file on your computer.

E-Utilities

Tools that provide access to data within NCBI's Entrez system outside of the regular web query interface. They provide a method of automating Entrez tasks within software applications. Each utility performs a specialized retrieval task, and can be used simply by writing a specially formatted URL.

Genome BLAST

This tool compares nucleotide or protein sequences to genomic sequence databases and calculates the statistical significance of matches using the Basic Local Alignment Search Tool (BLAST) algorithm.

Genome Remapping Service

NCBI's Remap tool allows users to project annotation data from one assembly to another through a base by base analysis. Options are provided to adjust the stringency of remapping, and summary results are displayed on the web page. Full results can be downloaded for viewing in NCBI's Genome Workbench graphical viewer, and annotation data for the remapped features, as well as summary data, is also available for download.

Genome Workbench

An integrated application for viewing and analyzing sequence data. With Genome Workbench, you can view data in publically available sequence databases at NCBI, and mix these data with your own data.

Open Reading Frame Finder (ORF Finder)

All | Databases | Downloads | Submissions | Tools | **How To**

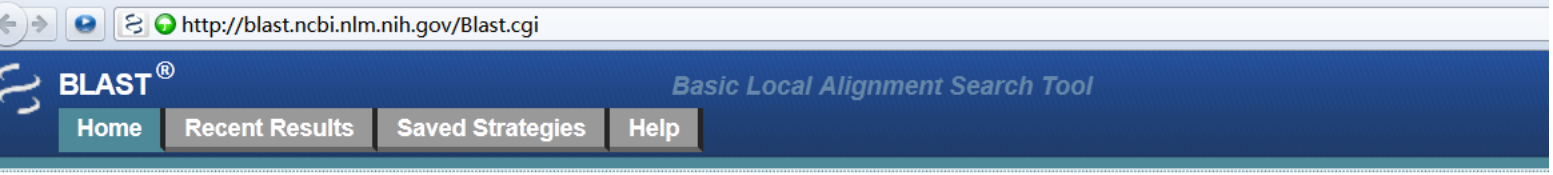
HOW TO

- View/download features around an object or between two objects on a chromosome
- Find a curated version of a sequence record (NCBI Reference Sequence)
- Find published information on a gene or sequence
- Find transcript sequences for a gene
- Link from an object on a map to another resource
- Design PCR primers and check them for specificity
- Retrieve all sequences for an organism or taxon
- Save text searches and set up automated searches with E-mailed results
- Submit data to NCBI
- Submit sequence data to NCBI
- Download a large, custom set of records from NCBI

Quick Links

- BLAST (Stand-alone) — 序列比对（单机版）
- Basic Local Alignment Search Tool (BLAST) — 序列比对（网络版）
- Cn3D — 三维结构可视化
- Conserved Domain Search Service (CD Search) — 保守结构域搜索
- E-Utilities — NCBI编程接口
- GenBank: BankIt
- GenBank: Sequin
- GenBank: tbl2asn

  NCBI序列提交工具 — 网页 / 图形 / 命令行
- Genome ProtMap — 蛋白质同源家族基因组映射
- Genome Workbench — 序列浏览和分析工具
- Primer-BLAST — PCR引物设计
- ProSplign — 蛋白质比对工具
- PubChem Structure Search — 化合物结构搜索
- SNP Submission Tool — SNP提交
- Splign — cDNA或剪切体向基因组比对工具
- Vector Alignment Search Tool (VAST) — 载体搜索工具



NCBI/ **BLAST Home**

BLAST finds regions of similarity between biological sequences. more...

New Aligning Multiple Protein Sequences? Try the COBALT Multiple Alignment Tool. Go

BLAST Assembled RefSeq Genomes

Choose a species genome to search, or list all genomic BLAST databases.

- Human
- Mouse
- Rat
- *Arabidopsis thaliana*
- *Oryza sativa*
- *Bos taurus*
- *Danio rerio*
- *Drosophila melanogaster*
- *Gallus gallus*
- *Pan troglodytes*
- *Microbes*
- *Apis mellifera*

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with Primer-BLAST
- Search trace archives
- Find conserved domains in your sequence (cds)
- Find sequences with similar conserved domain architecture (cdart)
- Search sequences that have gene expression profiles (GEO)
- Search immunoglobulins (IgBLAST)
- Search using SNP flanks
- Screen sequence for vector contamination (vecscreen)
- Align two (or more) sequences using BLAST (bl2seq)
- Search protein or nucleotide targets in PubChem BioAssay
- Search SRA transcript and genomic libraries
- Constraint Based Protein Multiple Alignment Tool
- Needleman-Wunsch Global Sequence Alignment Tool
- Search RefSeqGene
- Search WGS sequences grouped by organism

NCBI **Structure**

HOME | SEARCH | GUIDE | Structure Home | 3D Macromolecular Structures | Conserved Domains

Search Structure for GO CLEAR Help

3D Macromolecular Structures

RESOURCES

Resources

Molecular Modeling Database (MMDB)

Experimentally resolved structures of proteins, RNA, and DNA, derived from the Protein Data Bank (PDB), with value-added features such as explicit chemical graphs, computationally identified 3D domains (compact substructures) that are used to identify similar 3D structures, as well as links to literature, similar sequences, information about chemicals bound to the structures, and more. These connections make it possible, for example, to find 3D structures for homologs of a protein sequence of interest, then interactively view the sequence-structure relationships, active sites, bound chemicals, journal articles, and more.

Search | How To | Help | News | FTP | Publications

Retrieve by MMDB ID or PDB ID: Go

CBLAST

A tool that compares a query protein sequence against all protein sequences from resolved 3D structures by using protein BLAST against the PDB data set. The purpose is to find representative 3D structures for the query and/or its homologs, as available. Each record in the Entrez Protein database has been CBLAST'ed and the search results are available as Related Structures in the "Links" menu of Entrez Protein records. You can also enter a protein query sequence directly into the CBLAST search page in order to find its sequence-similar 3D structure records. The search results can be viewed in Cn3D (hence the name "CBLAST"), which displays an alignment of the query protein to the related structure's sequence and allows you to interactively examine the sequence-structure relationship.

Search | Help

Cn3D

A tool for visualization of three-dimensional structures with emphasis on interactive examination of sequence-structure relationships and superposition of geometrically similar structures. Can be used to display MMDB structures, superpositions of VAST related structures, and conserved core motifs identified in conserved domains.

About | Install | Tutorial | FAQ | Publications

IBIS

IBIS is the NCBI Inferred Biomolecular Interaction Server, which organizes, analyzes and predicts i…ins. IBIS provides annotations for different types of binding partners: protein, chemical, nucleic acid, peptide…entally determined structural complexes of a given protein, and at the same time infers binding sites/intera…y homologous proteins. Similar binding sites are clustered together based on their sequence and structure conserv…everal algorithms are used for verification in terms of evolutionary conservation, biological importance of l…s evidence from the published literature.

Search | Help | Publications

Vector Alignment Search Tool (VAST)

A computer algorithm developed at NCBI and used to identify similar protein 3-dimensional structu…omologs that cannot be recognized by sequence comparison. "Related structures" for every structure in MMDB ar…he MMDB Structure Summary pages. The VAST Search page also allows you to compare the coordinates of a …ures in MMDB to find its neighbors.

About | Search | VAST Help | VAST Search Help | FAQ | Linking | Publications

NCBI Resources How To

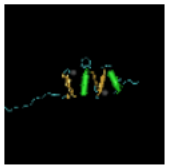
Structure

Structure ctcf

Save search Limits Advanced

Display Settings: Summary, Sorted by Default order

Results: 2

1. 

Solution Structure Of The Zinc Finger Domain Of Transcriptional Repressor Ctcf Protein[Transcription]

Taxonomy: Homo sapiens

Proteins: 1 **Chemicals:** 1 **modified:** 2011/02/17

MMDB ID: 36407 **PDB ID:** 2CT1

View in Cn3D PubChem Compound Protein

2. 

Solution Structures Of The C2h2 Type Zinc Finger Domain (

Taxonomy: Homo sapiens

Proteins: 1 **Chemicals:** 1 **modified:** 2011/02/17

MMDB ID: 35962 **PDB ID:** 1X6H

View in Cn3D PubChem Compound Protein

NCBI National Center for Biotechnology Information

Structure Summary

MMDB

PDB or MMDB ID Search

HOME SEARCH GUIDE Structure Home 3D Macromolecular Structures Conserved Domains Pubchem BioSystems Help

Solution Structure Of The Zinc Finger Domain Of Transcriptional Repressor Ctcf Protein

Author:
Structural GenomicsPROTEOMICS INITIATIVE (RSGI)
No PubMed record found.

MMDB ID: 36407 **PDB ID:** 2CT1

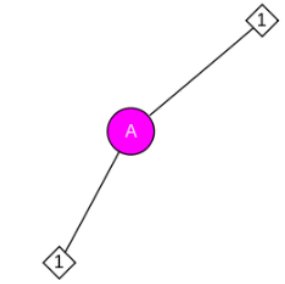
Deposited: 2005/5/23

Taxonomy: Homo sapiens

Related Structures: VAST

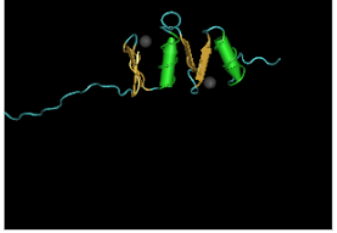
Experimental Method: NMR, 20 Structures

Interactions



Protein Nucleotide Chemical

Molecular Graphic



View or Save 3D Structure

File Format: Cn3D

Display As: 3D structure

Data Set: Single 3D structu

View structure

Download Cn3D

NOTICE

In order to view this biological unit properly, please upgrade to Cn3D 4.3.

Molecules and interactions

Label	Count	Molecule	Interactions with:
Protein			
A	1	Transcriptional Repressor Ctcf [redacted] Show sequence annotation ▼	• zinc ion
Chemicals			
1	2	Zn^{++} (zinc ion)	• Transcriptional Repressor Ctcf

* Click molecule labels to explore molecular sequence information.

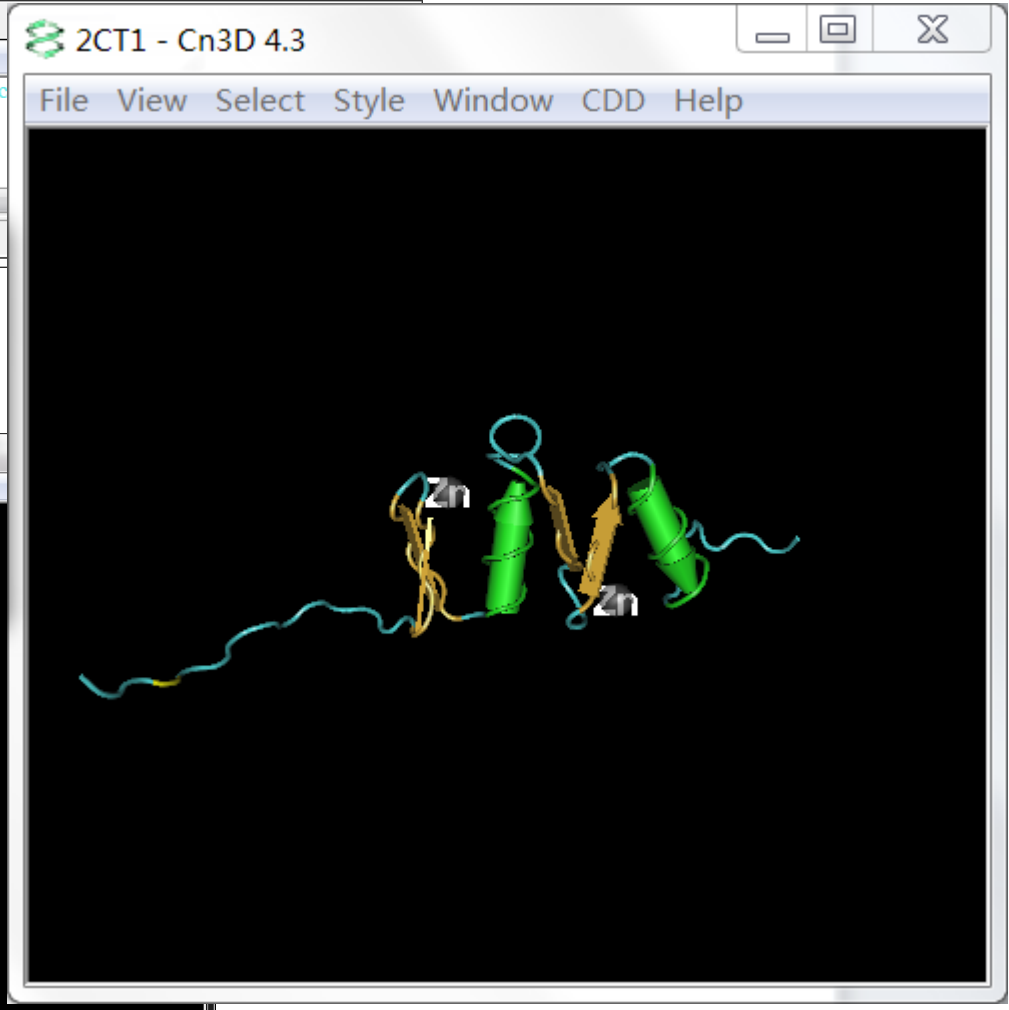
Citing MMDB

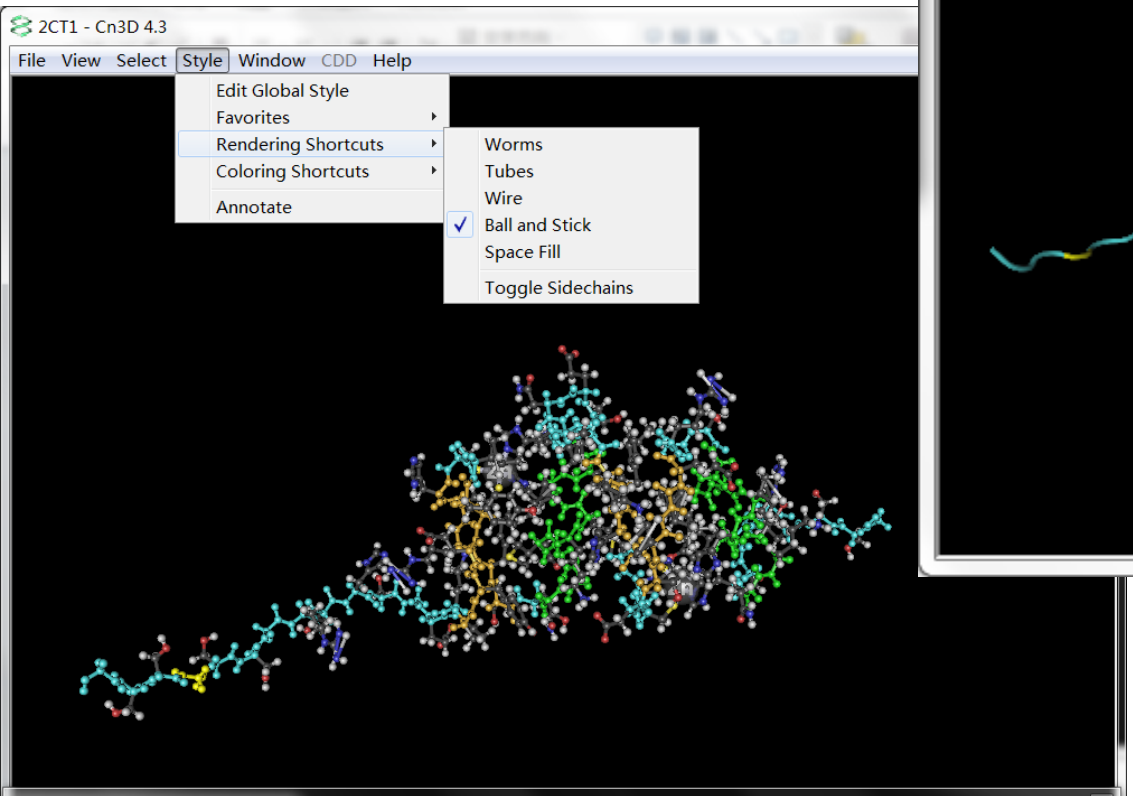
Wang Y, Addess KJ, Chen J, Geer LY, He J, He S, Lu S, Madej T, Marchler-Bauer A, Thiessen PA, Zhang N, Bryant SH. *"MMDB: annotating protein sequences with Entrez's 3D-structure database."*, **Nucleic Acids Res. 2007 Jan; 35(Database issue):** D298-300.

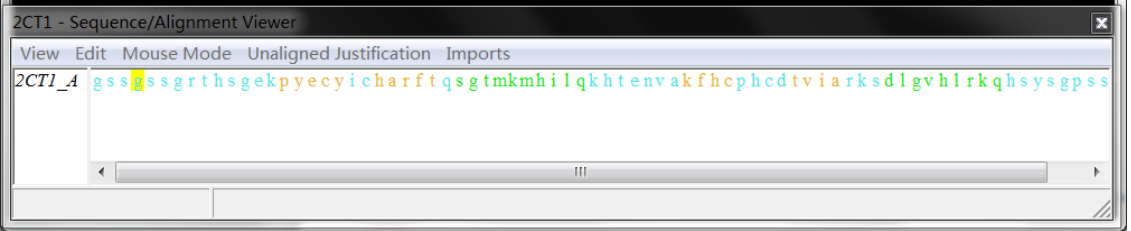
Bioinformatics

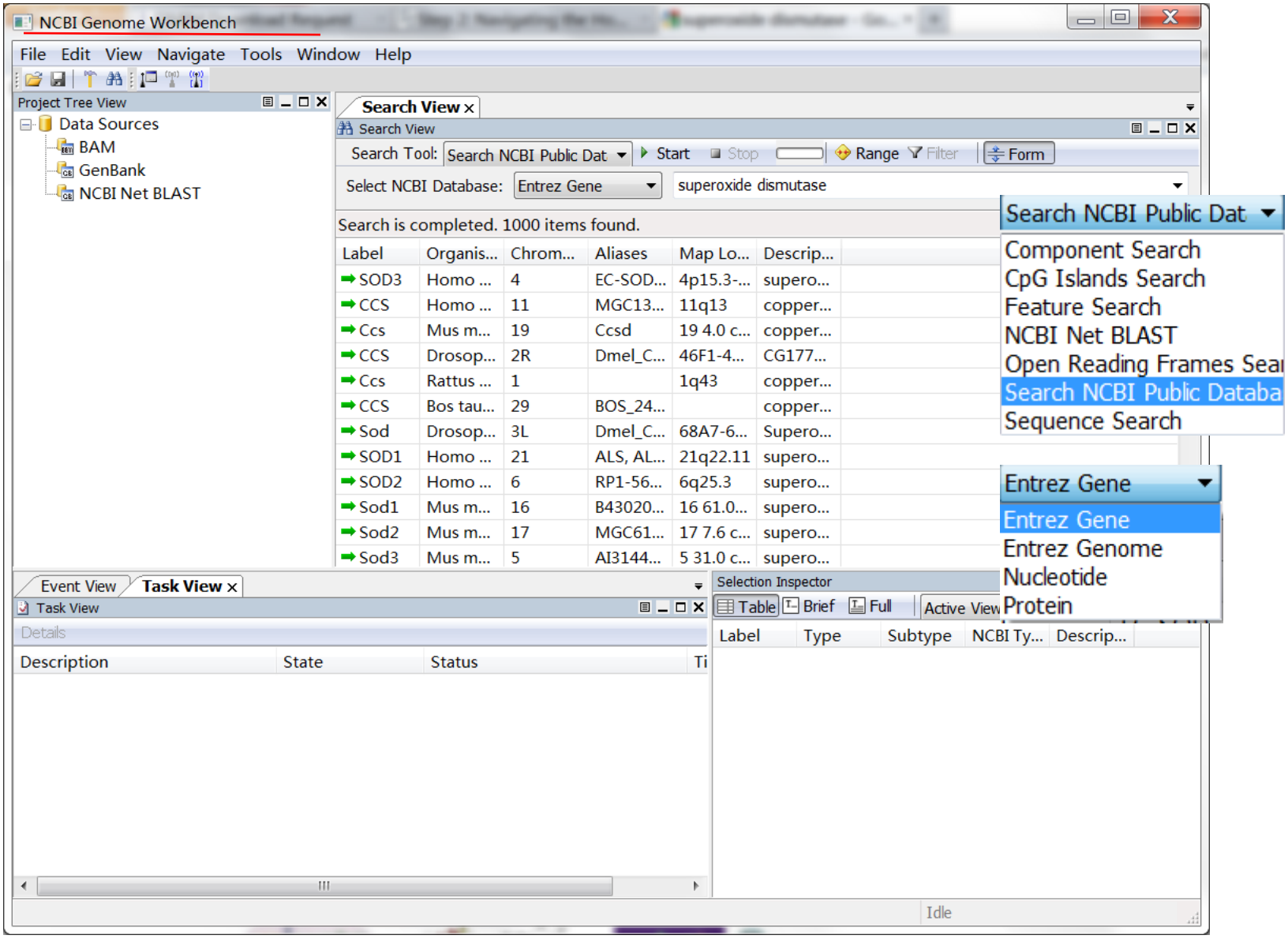
2011年8月


2CT1 - Sequence/Alignment Viewer
View Edit Mouse Mode Unaligned Justification Imports
2CT1_A
2CT1_A, loc 4 (PDB 4)

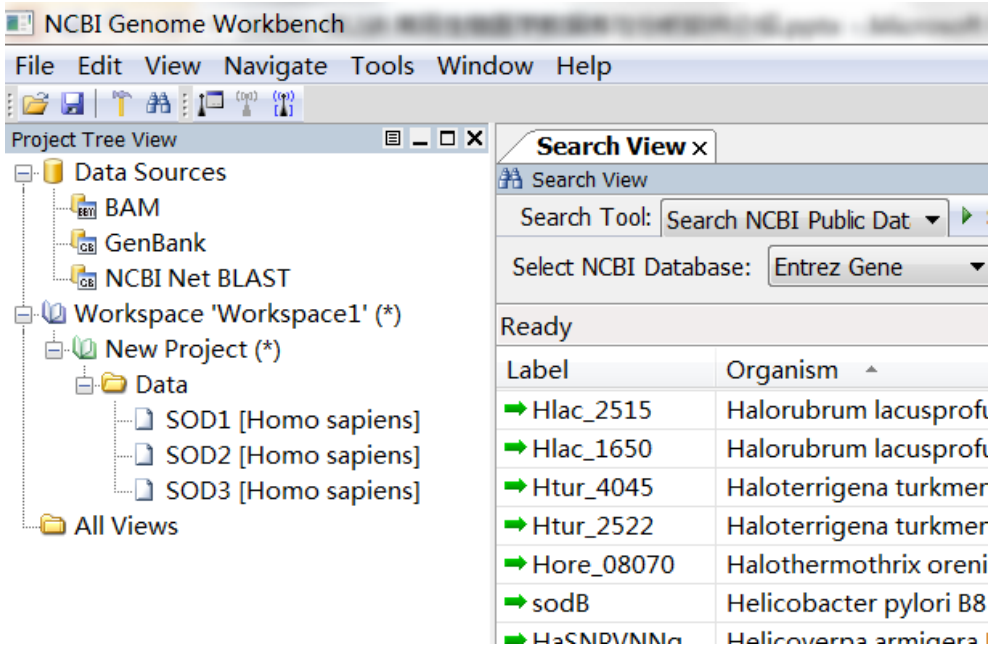

2CT1 - Cn3D 4.3
File View Select Style Window CDD Help
Zn
Zn

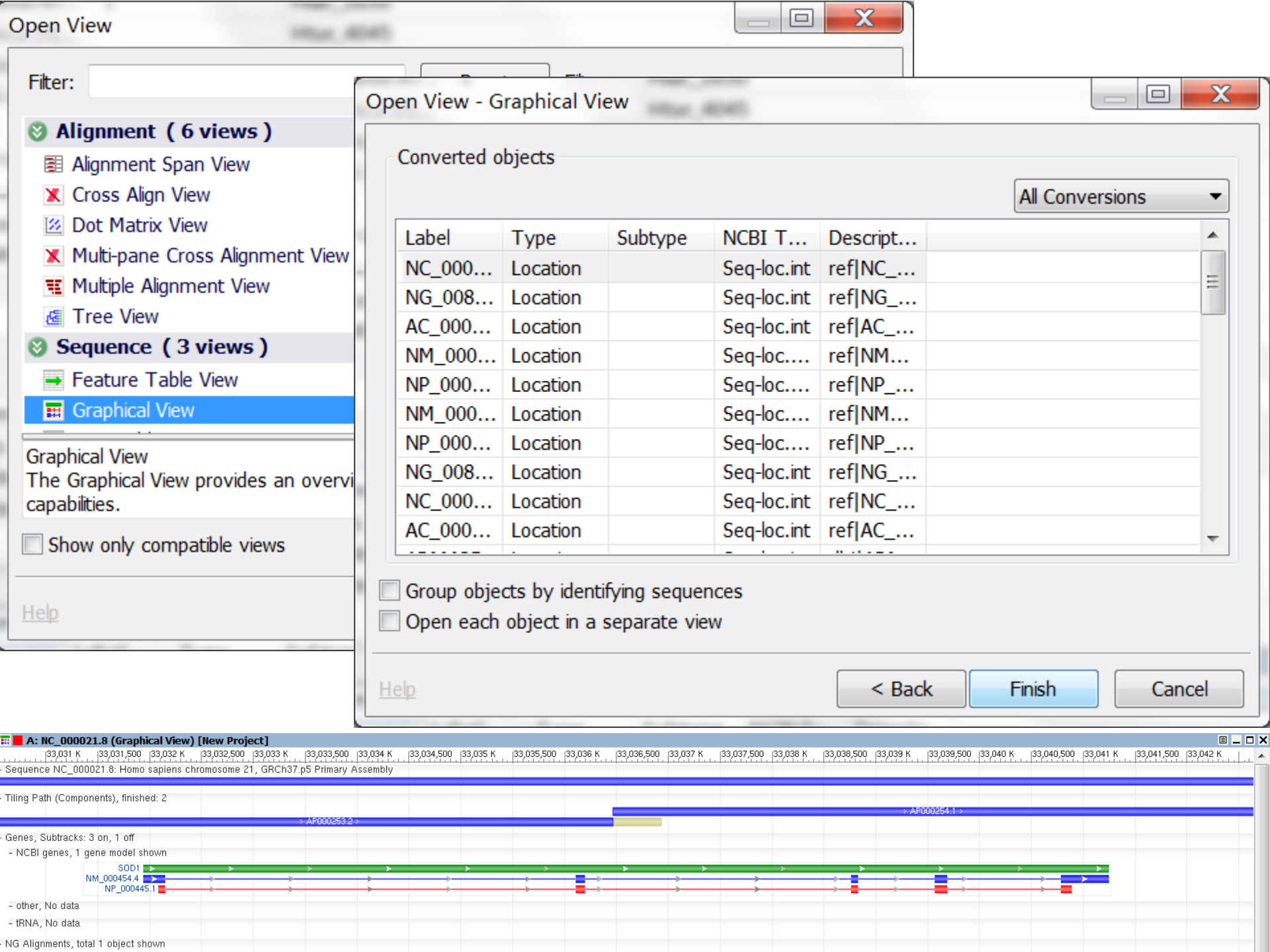

2CT1 - Cn3D 4.3
File View Select Style Window CDD Help
Edit Global Style
Favorites
Rendering Shortcuts
Coloring Shortcuts
Annotate
Worms
Tubes
Wire
Ball and Stick
Space Fill
Toggle Sidechains


2CT1 - Sequence/Alignment Viewer
View Edit Mouse Mode Unaligned Justification Imports
2CT1_A


NCBI Genome Workbench
File Edit View Navigate Tools Window Help
Project Tree View
Data Sources
BAM
GenBank
NCBI Net BLAST
Search View
Search Tool: Search NCBI Public Dat
Start
Stop
Range
Filter
Form
Select NCBI Database: Entrez Gene
superoxide dismutase
Search is completed. 1000 items found.
Label Organis... Chrom... Aliases Map Lo... Descrip...
SOD3 Homo ... 4 EC-SOD... 4p15.3-... supero...
CCS Homo ... 11 MGC13... 11q13 copper...
Ccs Mus m... 19 Ccsd 19 4.0 c... copper...
CCS Drosop... 2R Dmel_C... 46F1-4... CG177...
Ccs Rattus ... 1 1q43 copper...
CCS Bos tau... 29 BOS_24... copper...
Sod Drosop... 3L Dmel_C... 68A7-6... Supero...
SOD1 Homo ... 21 ALS, AL... 21q22.11 supero...
SOD2 Homo ... 6 RP1-56... 6q25.3 supero...
Sod1 Mus m... 16 B43020... 16 61.0... supero...
Sod2 Mus m... 17 MGC61... 17 7.6 c... supero...
Sod3 Mus m... 5 AI3144... 5 31.0 c... supero...
Search NCBI Public Dat
Component Search
CpG Islands Search
Feature Search
NCBI Net BLAST
Open Reading Frames Sear
Search NCBI Public Databa
Sequence Search
Entrez Gene
Entrez Gene
Entrez Genome
Nucleotide
Protein
Event View
Task View
Task View
Details
Description State Status Ti
Selection Inspector
Table Brief Full Active View
Label Type Subtype NCBI Ty... Descrip...
Idle

Label	Organism	Chromosome	Aliases	Map Location	Description
Hlac_2515	Halorubrum lacusprofundi AT...	1	Hlac_2515		manganese and iron superoxide dismutase
Hlac_1650	Halorubrum lacusprofundi AT...	1	Hlac_1650		manganese and iron superoxide dismutase
Htur_4045	Haloterrigena turkmenica DS...		Htur_4045		superoxide dismutase
Htur_2522	Haloterrigena turkmenica DS...		Htur_2522		superoxide dismutase
Hore_08070	Halothermothrix orenii H 168		Hore_08070		superoxide dismutase (Cu-Zn)
sodB	Helicobacter pylori B8		HPB8_439		superoxide dismutase, Fe-Mn family
HaSNPVNNg...	Helicoverpa armigera NPV N...		HaSNPVNNg1_gp109		superoxide dismutase
HaMNV_gp057	Helicoverpa armigera multipl...		HaMNV_gp057		ORF57; Sod
sodC	Herminiimonas arsenicoxydans		HEAR0840		superoxide dismutase (Cu-Zn)
Hbal_3001	Hirschia baltica ATCC 49814		Hbal_3001		superoxide dismutase Ni-type
SOD3	Homo sapiens	4	EC-SOD, MGC20077	4p15.3-p15.1	superoxide dismutase 3, extracellular
CCS	Homo sapiens	11	MGC138260	11q13	copper chaperone for superoxide dismutase
SOD1	Homo sapiens	21	ALS, ALS1, IPOA, SOD, hSod1, homodimer	21q22.11	superoxide dismutase 1, soluble
SOD2	Homo sapiens	6	RP1-56L9.2, IPOB, MNSOD, MVCD6	6q25.3	superoxide dismutase 2, mitochondrial
sodC	Hydrogenobacter thermophil...		HTH_0932		[Cu-Zn]
HY04AAS1_0...	Hydrogenobaculum sp. Y04A...		HY04AAS1_0075		Superoxide dismutase
Hden_1373	Hyphomicrobium denitrifican...		Hden_1373		superoxide dismutase copper/zinc binding protein
Hden_2549	Hyphomicrobium denitrifican...		Hden_2549		superoxide dismutase
sodc	Ictalurus punctatus				Cu-Zn superoxide dismutase




Open View
Filter:
Alignment (6 views)
Alignment Span View
Cross Align View
Dot Matrix View
Multi-pane Cross Alignment View
Multiple Alignment View
Tree View
Sequence (3 views)
Feature Table View
Graphical View
Graphical View
The Graphical View provides an overvi
capabilities.
Show only compatible views
Help
Open View - Graphical View
Converted objects
All Conversions
Label
Type
Subtype
NCBI T...
Descript...
NC_000...
Location
Seq-loc.int
ref|NC_...
NG_008...
Location
Seq-loc.int
ref|NG_...
AC_000...
Location
Seq-loc.int
ref|AC_...
NM_000...
Location
Seq-loc....
ref|NM...
NP_000...
Location
Seq-loc....
ref|NP_...
NM_000...
Location
Seq-loc....
ref|NM...
NP_000...
Location
Seq-loc....
ref|NP_...
NG_008...
Location
Seq-loc.int
ref|NG_...
NC_000...
Location
Seq-loc.int
ref|NC_...
AC_000...
Location
Seq-loc.int
ref|AC_...
Group objects by identifying sequences
Open each object in a separate view
Help
< Back
Finish
Cancel
A: NC_000021.8 (Graphical View) [New Project]
33,031 K
33,031,500
33,032 K
33,032,500
33,033 K
33,033,500
33,034 K
33,034,500
33,035 K
33,035,500
33,036 K
33,036,500
33,037 K
33,037,500
33,038 K
33,038,500
33,039 K
33,039,500
33,040 K
33,040,500
33,041 K
33,041,500
33,042 K
Sequence NC_000021.8: Homo sapiens chromosome 21, GRCh37.p5 Primary Assembly
Tiling Path (Components), finished: 2
AP000253.2
AP000254.1
Genes, Subtracks: 3 on, 1 off
NCBI genes, 1 gene model shown
SOD1
NM_000454.4
NP_000445.1
other, No data
tRNA, No data
NG Alignments, total 1 object shown

Welcome to Genome Workbench Tutorials!

Here you will find several examples and walk-throughs to guide you in how to use Genome Workbench to solve specific tasks. These tasks range from simple activities in learning how to use the interface to more complex examples of performing cross-species comparisons and comparing annotations on different sequences.

The available tutorials include:

Tutorial 1: Basic Operation

The first tutorial focuses on starting to work with Genome Workbench and on the basics of how the user interface works. This is done in by exploring a single gene on a large chromosome.

Tutorial 2: Working with Non-Public Data

Tutorial 2 shows how to manage projects and data in projects, particularly data that is not available at NCBI.

Tutorial 3: Working With Multiple Views

This tutorial describes how Genome Workbench integrates many different views together and how views can interact with one another. This tutorial also describes some of the more advanced [illegible]Workbench. This is done in the context of a study of cytochrome oxidase among *Collembola* species.

Tutorial 4: Genes and Variation

Variations describes ways in which Genome Workbench can be used to explore variations on genomic sequences.

Tutorial 5: Generating Sequence Overlap Alignments

Sequence alignments are of fundamental importance in analyzing sequence structures. This tutorial will describe several ways in which sequence alignments can be generated from withing Genome Workbench.

Tutorial 6: Working with BAM files

This exercise shows how to work with BAM files in gBench. It shows 4 different examples, starting with a sorted BAM file with index and converage graph, to a sorted BAM file with index and no coverage graph, then a sorted BAM file with no index and no coverage

Video Tutorials
Basic Features
Search for Genes
Move/Dock Windows
Phylogenetic Trees
Window Masker

Previous Next

Step 5: The Graphical View

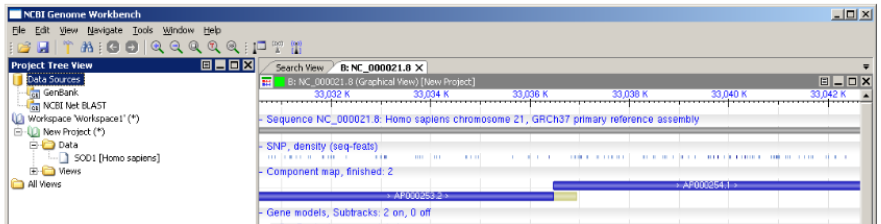
The graphical view shows the public annotations on a sequence, using both color and arrangement to show the relationships. In the view below, different annotations are shown with different colors:

- ***Green bars*** represent genes.
- ***Blue bars*** represent transcripts / mRNAs.
- ***Red bars*** represent coding regions / proteins.

At the bottom of the graphical view there are several controls. The content drop down is highlighted in the figure below and the SNP and the STS tracks are turned on.

In addition to the Central Dogma annotations, there are additional annotations that are available. These include:

- At the top of the image is a row of short blue tick marks. These represent ***variations*** from dbSNP for this sequence. As you zoom in and out, these will become available as selectable things.
- Just beneath the variations are a set of blue bars representing the ***components*** that are used to assemble this sequence. Most large genomic sequences are split into many smaller pieces and reassembled from these chunks; the blue bars show you where the chunk boundaries are, and what the approximate overlap between chunks is.
- Many other features, including ***sequence tagged sites*** (STSs, visible in the image below) are shown as black bars underneath the genes and gene products.



Search: All species for Go

200000 or coronary heart disease

All species

Favourite species
Human
Mouse
Zebrafish

Alpaca
Anole Lizard
Armadillo
Bushbaby
C. elegans
C. intestinalis
C. savignyi
Cat
Chicken
Chimpanzee
Cow
Dog
Dolphin

Browse a Genome

The Ensembl project produces genome (eukaryotic species, and makes this information freely available online.

Click on a link below to go to the species

Popular genomes (Log in to custom

Human
GRCh37

Mouse
NCBIM37

Zebrafish
Zv9

All genomes

-- Select a species --

View full list of all Ensembl species

Other species are available in Ensembl *Pre!* and EnsemblGenomes

Ensembl is a joint project between EMBL - EBI and the Wellcome Trust Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.

Ensembl receives major funding from the Wellcome Trust. Our acknowledgements page includes a list of additional current and previous funding bodies.

EMBL-EBI

New to Ensembl?

Did you know you can:

- **Learn how to use Ensembl** with our video tutorials and walk-throughs
- **Add custom tracks** using our new Control Panel
- **Upload and analyse your data** and save it to your Ensembl account
- **Search for a DNA or protein sequence** using BLAST or BLAT
- **Fetch only the data you want** from our public database, using the Perl API
- **Download our databases via FTP** in FASTA, MySQL and other formats
- **Mine Ensembl with BioMart** and export sequences or tables in text, html, or Excel format

Still got questions? Try our FAQs or glossary

What's New in Release 63 (30 June 2011)

- Sortable tracks on Region in Detail
- Variant Effect Predictor 2.1
- Regulation configuration matrix redesign

Full details of this release

More release news on our blog →

Latest blog posts

- Run a private Ensembl MySQL in the cloud
- More Ensembl vacancies
- New features in the Variant Effect Predictor

Go to Ensembl blog →

You've been redirected to your nearest mirror - asia.ensembl.org

- Take me back to www.ensembl.org

Download a sequence or region

Click on the 'Export data' button in the lefthand menu of most pages to export:

- FASTA sequence
- GTF or GFF features

...and more!

Customise your download

Custom datasets can be retrieved using the BioMart data-mining tool.

You may find exploring this web-based query tool easier than extracting information direct from our databases.

Fetch data programmatically

Write your own Perl scripts to retrieve small-to-medium datasets. All our data, as well as added functionality, is available through the Ensembl Perl API.

Use the API to retrieve gene and transcript sets, fetch alignments between sequences, compare allele frequencies and much more!

Download databases & software

All of our data and software, including pipelines and web code, is available free.

- Download data via FTP
- Ensembl pipeline in CVS
- Set up your own Ensembl website

Export via website

Many of the pages displaying Ensembl genomic data offer an export option, suitable for small amounts of data.

This is the ideal option if you want a short sequence as a FASTA file, or a GFF file of a few features in a region. Simply find your desired feature or region, click on one of the "Export..." links in the left-hand menu, and select your output options.

We respectfully request that you do not script against the export pages on the Ensembl website, as this degrades the service for other web visitors. The public MySQL server is provided specifically for this purpose. Thank you.

If you wish to extract multiple features or regions, we recommend using the Perl API if possible.

The Ensembl public MySQL Servers

For large amounts of data and more detailed analysis, we recommend you use our publicly-accessible MySQL server, **ensembldb.ensembl.org**, which you can access as user 'anonymous'. A second server, **martdb.ensembl.org** provides public access to the BioMart databases.

There are two options for retrieving data from the Ensembl databases:

Perl API

The easiest option is to use the Ensembl Perl API. The API uses an object-oriented approach to model real biological objects such as genes, transcripts and 'slices' of DNA sequence, making it straightforward for you to write scripts that retrieve and analyze data.

```
# find gene stable ids corresponding to the MGI symbol 'Tyr' my
$gene_adaptor = $dbCore->get_GeneAdaptor(); my @genes = @{
$gene_adaptor->fetch_all_by_external_name('Tyr') };
```

Additional code examples and suggestions can be found in the various API tutorials and the ensembl-dev mailing list archives.

MySQL client

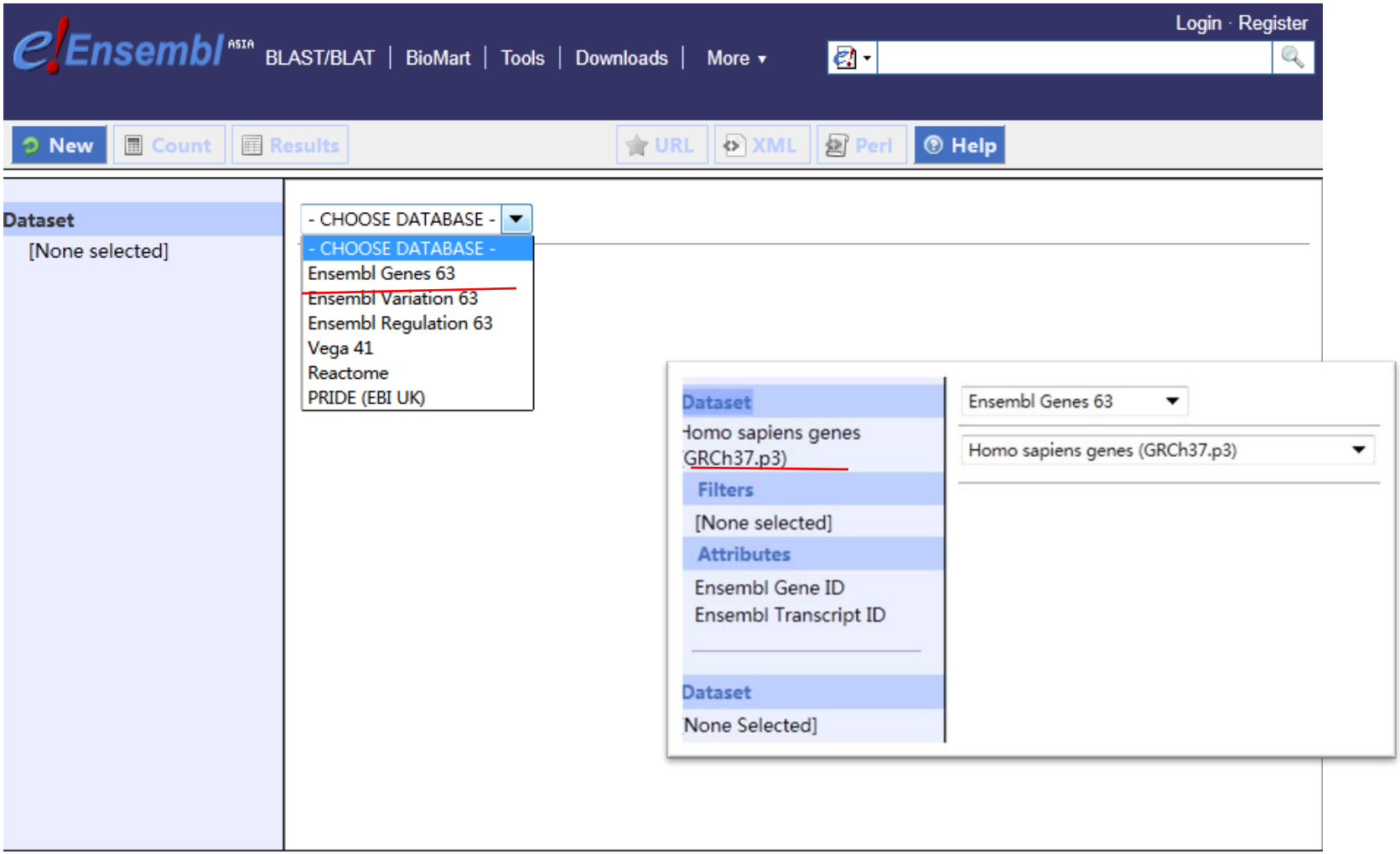
Alternatively, you can use a MySQL client program to query the database directly. However this does require knowledge of the complex database schemas used by Ensembl, so we only recommend it if you are unable to install our API (e.g. if you have no access to Perl). Also note that direct MySQL queries on the database are not suited to retrieve sequences. To retrieve sequences please use the Perl API.

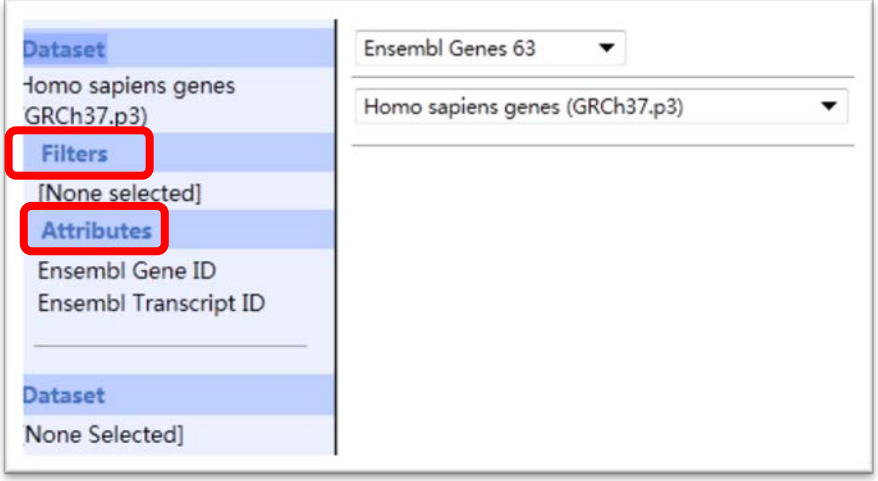
For **BioMart** access, we strongly recommend that you use the martview web interface, as the mart databases contain very many tables of denormalised data. Data can also be retrieved from BioMart programmatically, using the Mart XML-based webservice.

Show All entries | Show/hide columns | Filter

Species	DNA (FASTA)	cDNA (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (EMF)	Variation (GVF)	Regulation (GFF)	Data files
Ailuropoda melanoleuca (Panda)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Anolis carolinensis (Anole Lizard)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Bos taurus (Cow)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	GVF	-	-
Caenorhabditis elegans (C.elegans)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Callithrix jacchus (Marmoset)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Canis familiaris (Dog)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	GVF	-	-
Cavia porcellus (Guinea Pig)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Choloepus hoffmanni (Sloth)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Ciona intestinalis (C.intestinalis)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Ciona savignyi (C.savignyi)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Danio rerio (Zebrafish)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	GVF	-	-
Dasypus novemcinctus (Armadillo)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Dipodomys ordii (Kangaroo rat)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Drosophila melanogaster (Fruitfly)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	GVF	-	-
Echinops telfairi (Lesser hedgehog tenrec)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Equus caballus (Horse)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	GVF	-	-
Erinaceus europaeus (Hedgehog)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	-	-
Felis catus (Cat)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	GVF	-	-
Gallus gallus (Chicken)	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	GVF	-	-

Ensembl: BioMart





Please restrict your query using criteria below

- ⊞ REGION:
- ⊟ GENE:
  - Limit to genes ... with Illumina HumanWG 6 v1 ID(s)) — Only / Excluded
  - ID list limit — Ensembl Gene ID(s) [e.g. ENSG00000139618] — 浏览…
  - Transcript count >=
  - Gene type: IG_C_gene, IG_C_pseudogene, IG_D_gene, IG_J_gene, IG_J_pseudogene, IG_V_gene
  - Source: ensembl
  - Status (gene): KNOWN
  - Status (transcript): KNOWN
- ⊞ TRANSCRIPT EVENT:
- ⊞ GENE ONTOLOGY:
- ⊞ EXPRESSION:
- ⊞ MULTI SPECIES COMPARISONS:
- ⊞ PROTEIN DOMAINS:
- ⊞ VARIATION:

Please select columns to be included in the output and hit 'Results' when ready

- **Features** / **Homologs**
- **Structures** / **Variation**
- **Transcript Event** / **Sequences**

⊟ GENE:

Ensembl

- [x] Ensembl Gene ID
- [x] Ensembl Transcript ID
- [ ] Ensembl Protein ID
- [ ] Canonical transcript stable ID(s)
- [ ] Description
- [ ] Chromosome Name
- [ ] Gene Start (bp)
- [ ] Gene End (bp)
- [ ] Strand
- [ ] Band
- [ ] Transcript Start (bp)
- [ ] Transcript End (bp)
- [ ] Associated Gene Name
- [ ] Associated Transcript Name
- [ ] Associated Gene DB
- [ ] Associated Transcript DB
- [ ] Transcript count
- [ ] % GC content
- [ ] Gene Biotype
- [ ] Transcript Biotype
- [ ] Source
- [ ] Status (gene)
- [ ] Status (transcript)

- ⊞ EXTERNAL:
- ⊞ EXPRESSION:
- ⊞ PROTEIN DOMAINS:

UCSC

UCSC Genome Bioinformatics

Genomes - Blat - Tables - Gene Sorter - PCR - VisiGene - Proteome - Session - FAQ - Help

Genome Browser
ENCODE
Neandertal
Blat
Table Browser
Gene Sorter
In Silico PCR
Genome Graphs
Galaxy
VisiGene
Proteome Browser
Utilities
Downloads
Release Log
Custom Tracks
Microbial Genomes
Mirrors
Archives
Training
Credits
Publications
Cite Us
Licenses
Jobs
Staff
Contact Us

About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to the ENCODE and Neandertal projects.

We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, homology and other information on groups of genes that can be related in many ways. Blat quickly maps your sequence to the genome. The Table Browser provides convenient access to the underlying database. VisiGene lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. Genome Graphs allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE) at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our public mailing list.

News News Archives ▶

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the genome-announce mailing list.

`16 June 2011 - Re-engineered OMIM Tracks Released`

We announce today the release of our newly re-engineered OMIM (Online Mendelian Inheritance in Man) tracks for both hg18 and hg19. With the kind assistance of Ada Hamosh (director), Joanna Amberger and Francois Schiettecatte of the OMIM project, we have divided the OMIM records into three separate tracks:

OMIM Allelic Variant SNPs
Variants in the OMIM database that have associated dbSNP identifiers.

OMIM Genes
The genomic positions of gene entries in the OMIM database. The coloring indicates the associated OMIM phenotype class.

OMIM Phenotypes - Gene Unknown
Regions known to be associated with a phenotype, but for which no specific gene is known to be causative. This track also includes known multi-gene syndromes.

The OMIM tracks are searchable by OMIM number. In most cases, simply typing the 6-digit MIM number into the position/search box on the Browser will take you to the record.

The OMIM data are the property of Johns Hopkins University and will not be available for download from UCSC. Please contact the OMIM project at omim.org for download information.

UCSC thanks engineers Fan Hsu, Brooke Rhead and Robert Kuhn for this release.

`9 June 2011 - UCSC Preview Browser Available`

Early access to ENCODE and other UCSC browser data tracks under construction is now available from the new UCSC Preview Browser site: http://genome-preview.ucsc.edu. Read more.

`7 June 2011 - Updated Lizard Browser Available:` We have released a Genome Browser for the May 2010 genome assembly of the green anole lizard, *Anolis carolinensis* (Broad version AnoCar2.0, UCSC version anoCar2). Read more.

`26 May 2011 - New Release of UCSC Genes for Mouse:` We've released an updated set of UCSC Genes for the mm9 (NCBI Build 37) mouse Genome Browser. Read more.

==> News Archives

Conditions of Use

The sequence and annotation data displayed in the Genome Browser are freely available for any use with the following conditions:

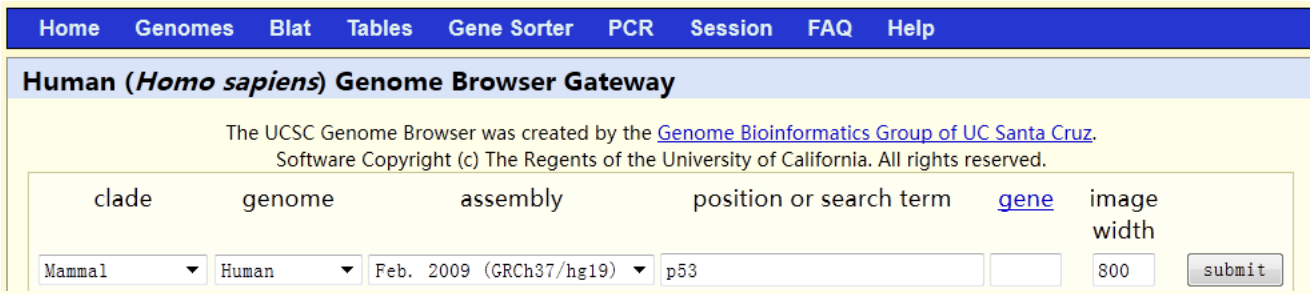
- Genome sequence data use restrictions are noted within the species sections on the Credits page.
- Some annotation tracks contributed by external collaborators contain proprietary data that have specific use restrictions. To check for restrictions associated with a particular genome assembly, review the *database/README.txt* file in the assembly's downloads directory.

The UCSC, Ensembl, and NCBI browser and annotation groups have established a common set of minimum requirements for public display of genome data made available after Spring 2009, described here.

The Genome Browser and Blat software are free for academic, nonprofit, and personal use. A license is required for commercial use. See the Licenses page for more information.

Program-driven use of this software is limited to a maximum of one hit every 15 seconds and no more than 5,000 hits per day.

UCSC: Genome Browser



Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061;RH80175	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
AA205474	Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101	Displays region of clone with GenBank accession AC008101
AF083811	Displays region of mRNA with GenBank accession number AF083811
PRNP	Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414	Displays the region of genome with RefSeq identifier NM_017414
NP_059110	Displays the region of genome with protein accession number NP_059110
pseudogene mRNA	Lists transcribed pseudogenes, but not cDNAs
homeobox caudal	Lists mRNAs for caudal homeobox genes
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
zahler	Lists mRNAs deposited by scientist named Zahler
Evans,J.E.	Lists mRNAs deposited by co-author J.E. Evans

Use this last format for author queries. Although GenBank requires the search format *Evans JE*, internally it uses the format *Evans,J.E.*.

Error(s):

- Search terms are not very specific, only showing first 500 matching known genes.

OK

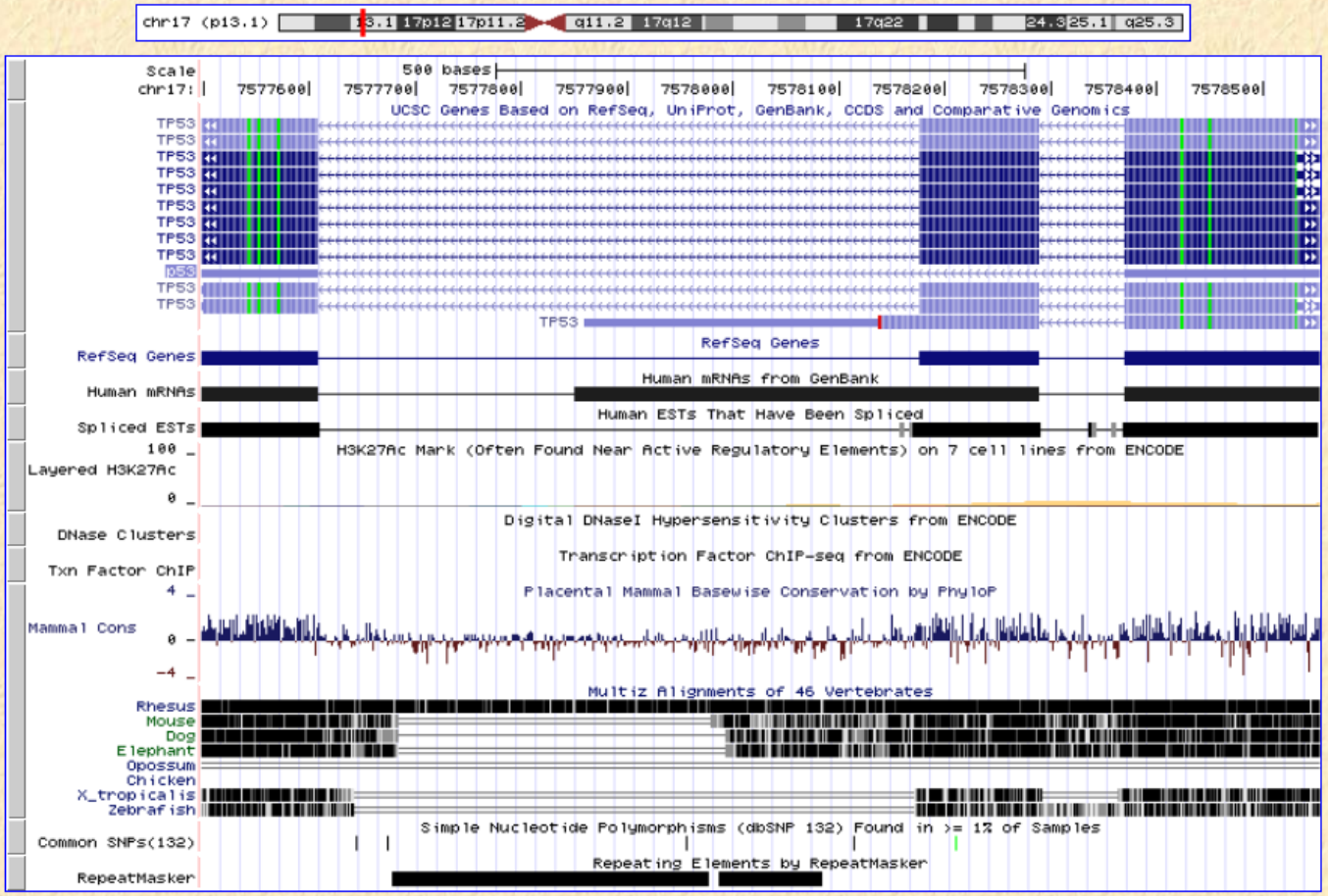
UCSC Genes

p53 (uc010cnj.1) at chr17:7577499-7578554 - Homo sapiens mRNA for P53, complete cds.
p53 (uc010cne.1) at chr17:7571720-7576926 - Homo sapiens mRNA for P53, complete cds.
ZMAT3 (uc010hxa.2) at chr3:178741527-178789584 - p53 target zinc finger protein isoform 2
CUL9 (uc003oul.2) at chr6:43149922-43192324 - p53-associated parkin-like cytoplasmic protein
CUL9 (uc003ouk.2) at chr6:43149922-43192324 - p53-associated parkin-like cytoplasmic protein
ZMAT3 (uc003fji.2) at chr3:178741527-178789584 - p53 target zinc finger protein isoform 2
ZMAT3 (uc003fjg.2) at chr3:178741527-178789584 - p53 target zinc finger protein isoform 1
TP53RK (uc002xsk.2) at chr20:45313005-45318276 - p53-related protein kinase
PDRG1 (uc002wxd.2) at chr20:30532759-30539883 - p53 and DNA damage-regulated protein
TRIAP1 (uc001tyg.2) at chr12:120881765-120884215 - p53-inducible cell-survival factor
TP53I11 (uc001myl.2) at chr11:44953900-44972608 - p53-induced protein
TP53I11 (uc001myk.2) at chr11:44953900-44972608 - p53-induced protein
TP53I11 (uc001myj.2) at chr11:44953900-44972608 - p53-induced protein
TP53I11 (uc001myi.2) at chr11:44953900-44971711 - p53-induced protein
PRG1 (uc010xwn.1) at chr19:43853208-43853700 - Homo sapiens p53-responsive gene 1 (PRG1), non-coding RNA.

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr17:7,577,499-7,578,554 gene jump clear size 1,056 bp. configure



move start < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks.

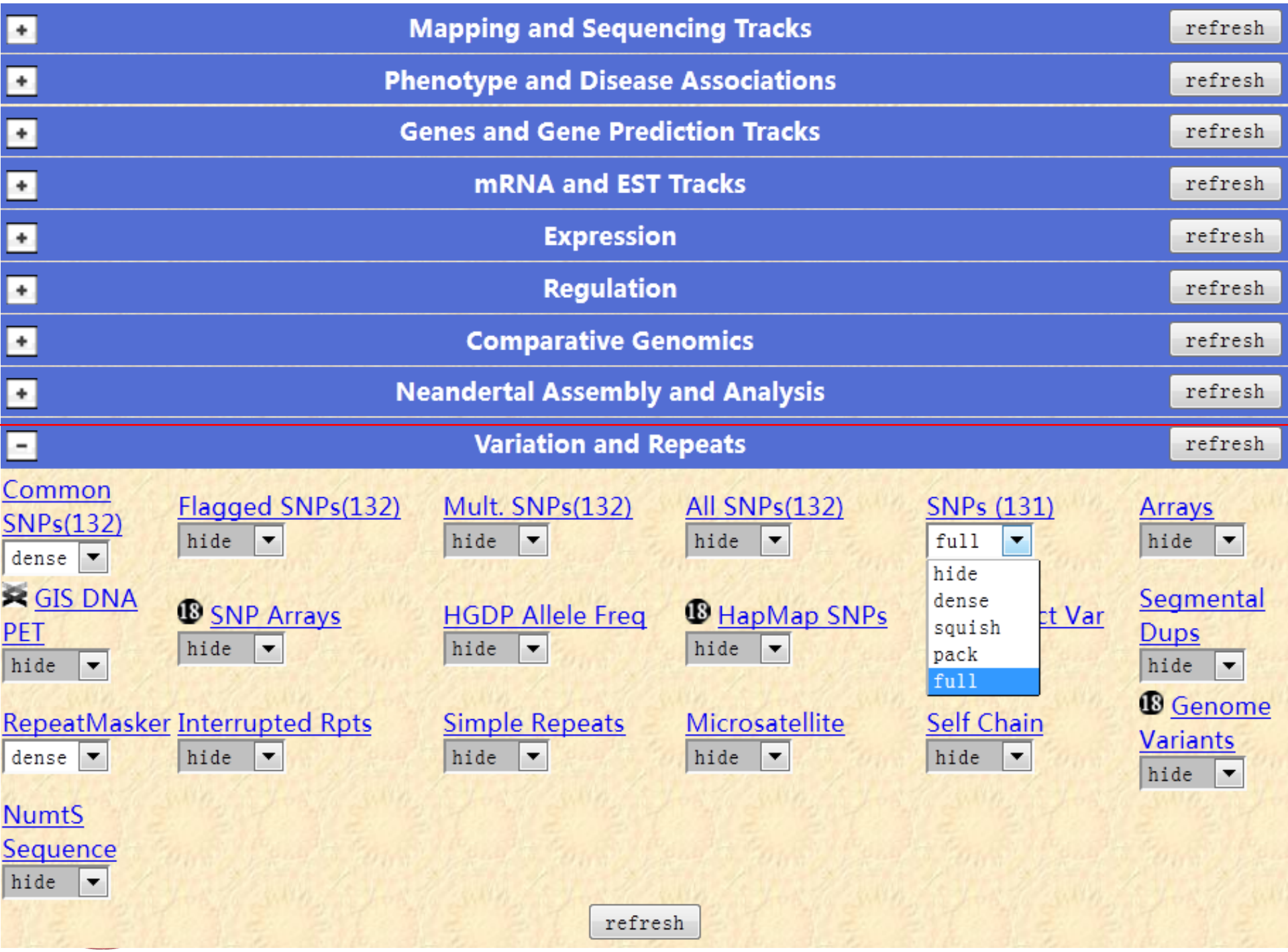
move end < 2.0 >

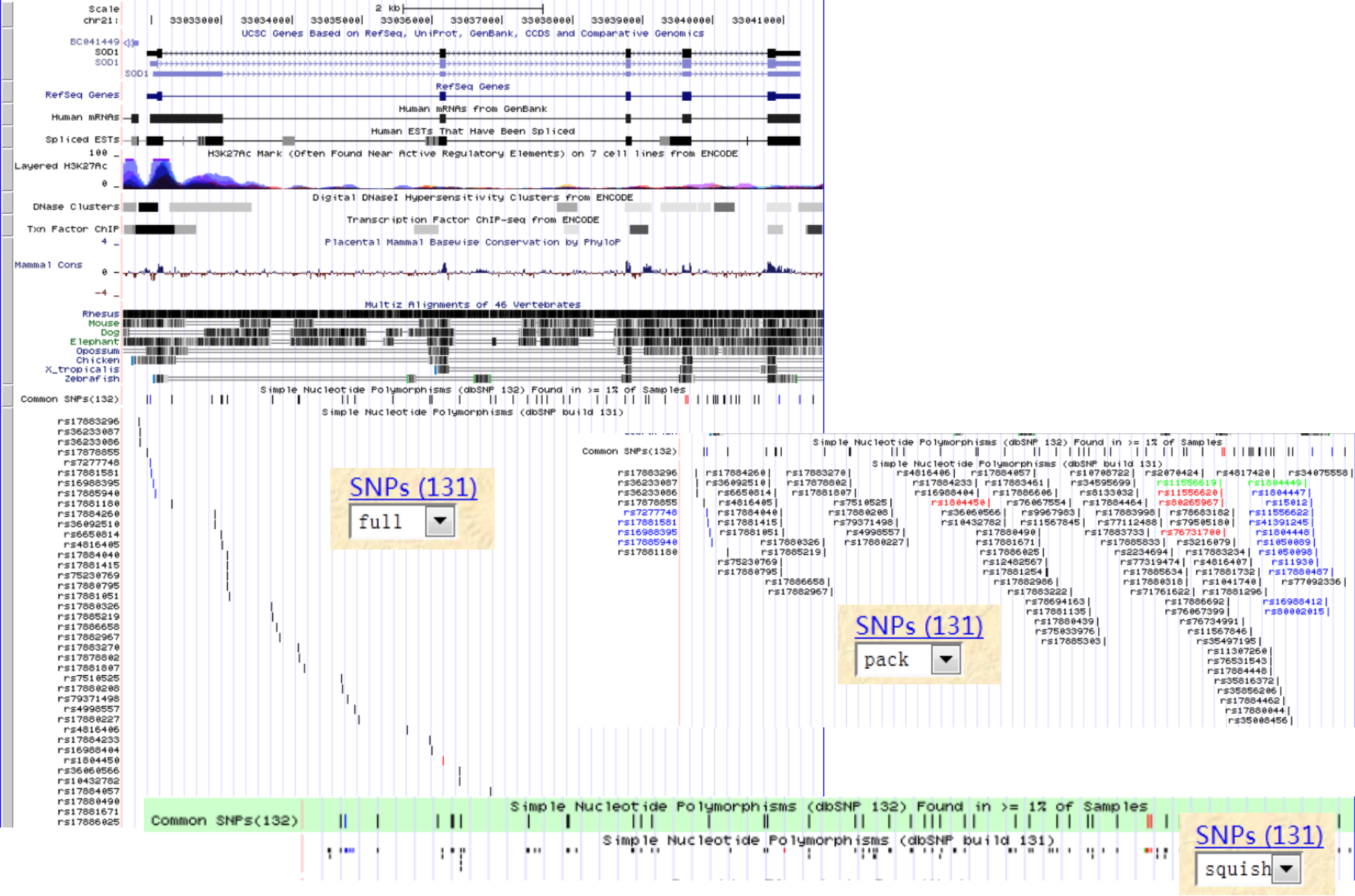
track search default tracks default order hide all add custom tracks configure reverse refresh

collapse all

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

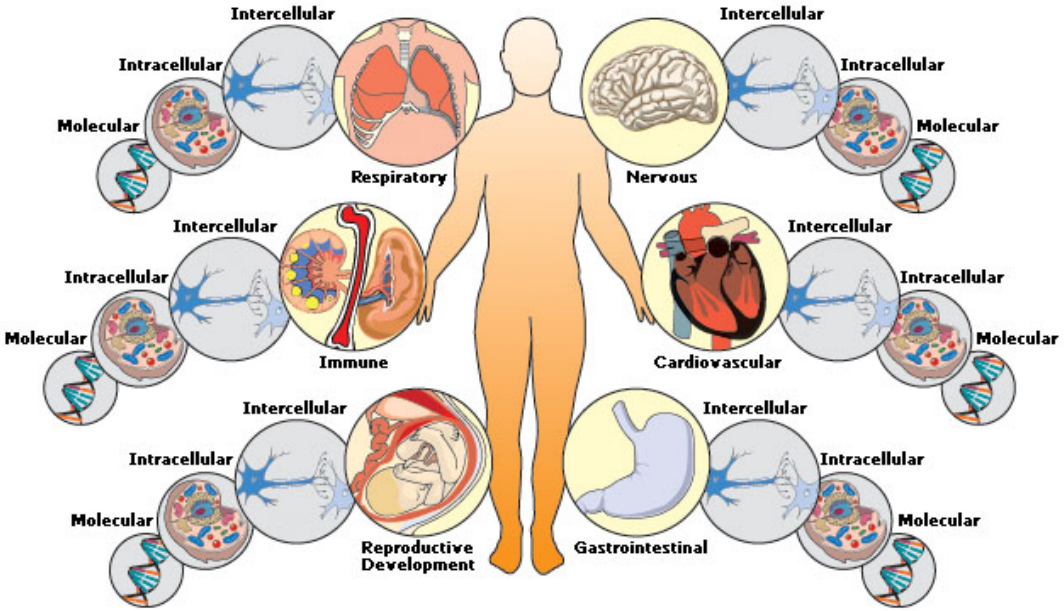
expand all


Mapping and Sequencing Tracks
refresh
Phenotype and Disease Associations
refresh
Genes and Gene Prediction Tracks
refresh
mRNA and EST Tracks
refresh
Expression
refresh
Regulation
refresh
Comparative Genomics
refresh
Neandertal Assembly and Analysis
refresh
Variation and Repeats
refresh
Common SNPs(132)
dense
Flagged SNPs(132)
hide
Mult. SNPs(132)
hide
All SNPs(132)
hide
SNPs (131)
full
hide
dense
squish
pack
full
Arrays
hide
GIS DNA PET
hide
SNP Arrays
hide
HGDP Allele Freq
hide
HapMap SNPs
hide
Segmental Dups
hide
RepeatMasker
dense
Interrupted Rpts
hide
Simple Repeats
hide
Microsatellite
hide
Self Chain
hide
Genome Variants
hide
NumtS Sequence
hide
refresh


Scale
chr21:
33033000
33034000
33035000
33036000
33037000
33038000
33039000
33040000
33041000
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
BC041449
SOD1
RefSeq Genes
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Digital DNaseI Hypersensitivity Clusters from ENCODE
Transcription Factor ChIP-seq from ENCODE
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 46 Vertebrates
Rhesus
Mouse
Dog
Elephant
Opossum
Chicken
X_tropicalis
Zebrafish
Simple Nucleotide Polymorphisms (dbSNP 132) Found in >= 1% of Samples
Simple Nucleotide Polymorphisms (dbSNP build 131)
Common SNPs(132)
SNPs (131)
full
pack
squish

复杂疾病数据库

- OMIM、GAD、CGAP等复杂疾病数据库主要基于文献、关联分析及生物学实验的结果，记录了疾病表型、相关的染色体区域、候选基因等多方面的信息。

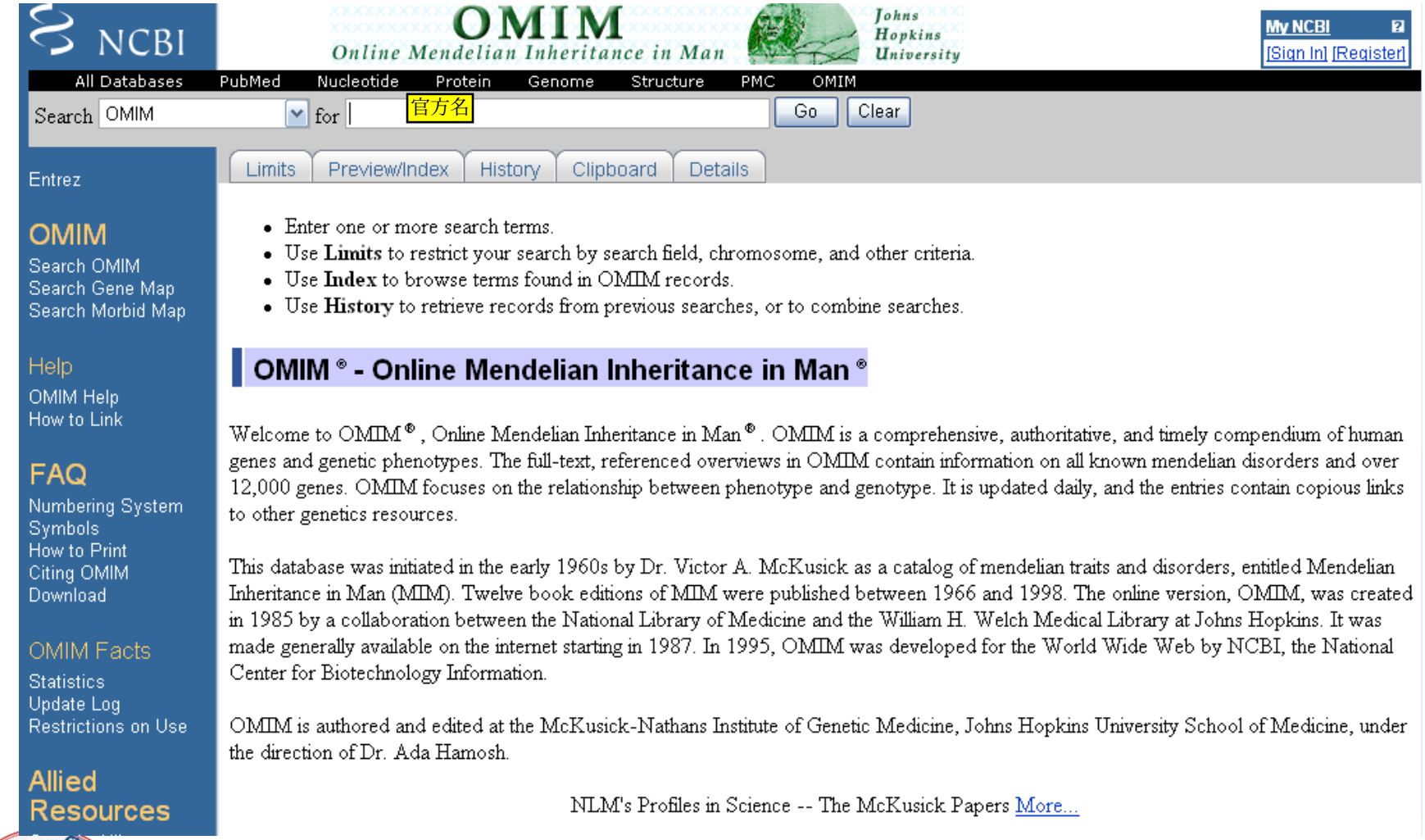


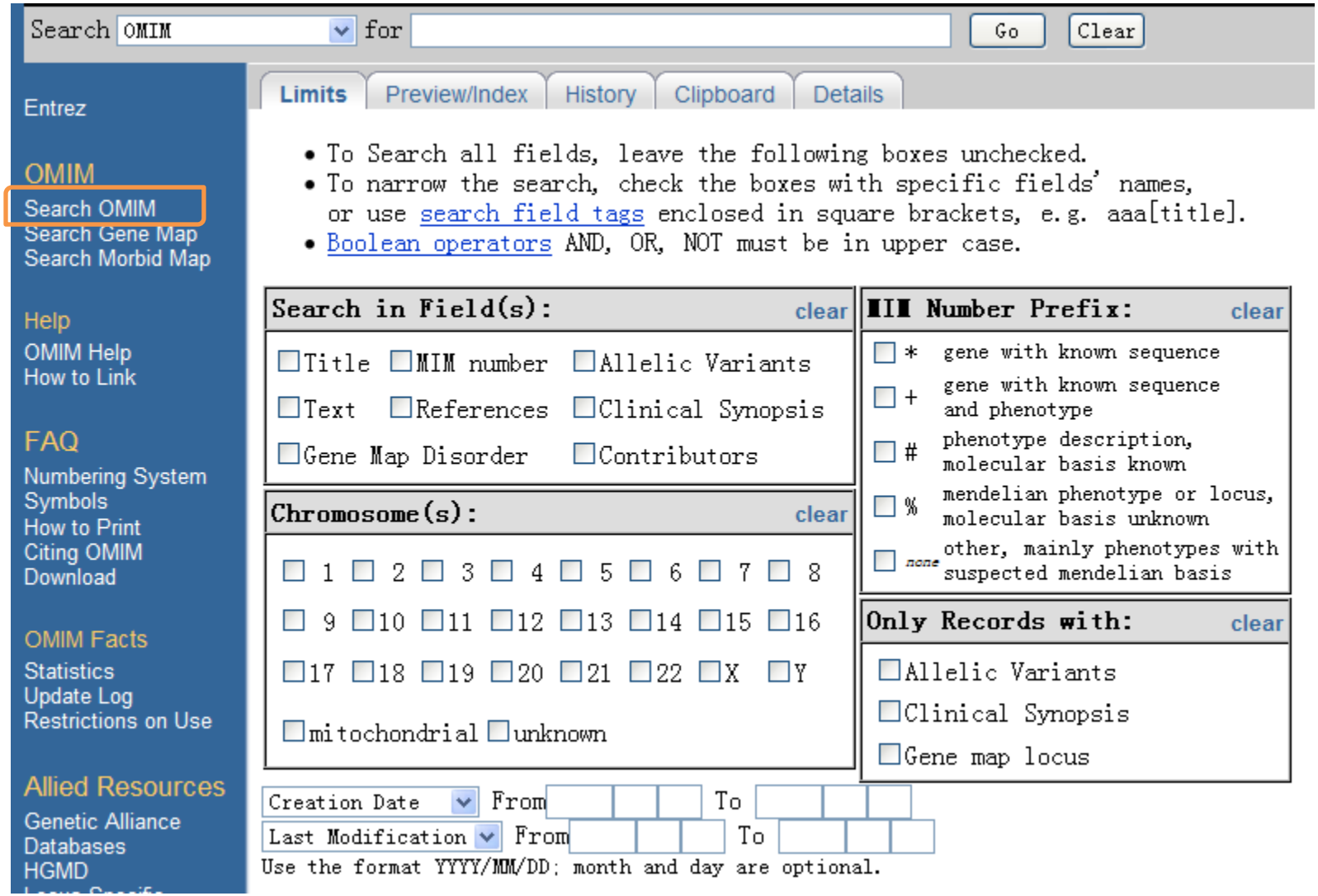
7k多疾病信

人类孟德尔遗传在线（OMIM）

- MIM (Mendelian Inheritance in Man)与OMIM
- OMIM的发展史
- 目前，OMIM的发布以及相关软件的开发由National Center for Biotechnology Information (NCBI)负责。
- OMIM 数据库的访问 http://www.ncbi.nlm.nih.gov/omim

OMIM 主页(http://www.ncbi.nlm.nih.gov/omim)




Search OMIM for Go Clear
Entrez
OMIM
Search OMIM
Search Gene Map
Search Morbid Map
Help
OMIM Help
How to Link
FAQ
Numbering System
Symbols
How to Print
Citing OMIM
Download
OMIM Facts
Statistics
Update Log
Restrictions on Use
Allied Resources
Genetic Alliance
Databases
HGMD
Limits Preview/Index History Clipboard Details
To Search all fields, leave the following boxes unchecked.
To narrow the search, check the boxes with specific fields' names, or use search field tags enclosed in square brackets, e.g. aaa[title].
Boolean operators AND, OR, NOT must be in upper case.
Search in Field(s): clear
Title MIM number Allelic Variants
Text References Clinical Synopsis
Gene Map Disorder Contributors
Chromosome(s): clear
1 2 3 4 5 6 7 8
9 10 11 12 13 14 15 16
17 18 19 20 21 22 X Y
mitochondrial unknown
MIM Number Prefix: clear
* gene with known sequence
+ gene with known sequence and phenotype
# phenotype description, molecular basis known
% mendelian phenotype or locus, molecular basis unknown
none other, mainly phenotypes with suspected mendelian basis
Only Records with: clear
Allelic Variants
Clinical Synopsis
Gene map locus
Creation Date From To
Last Modification From To
Use the format YYYY/MM/DD; month and day are optional.

Search OMIM for Alzheimer's disease Go Clear Save Search

Limits | Preview/Index | History | Clipboard | Details

Display Titles Show 20 Send to

All: 150 | OMIM UniSTS: 32 | OMIM dbSNP: 59

Items 1 - 20 of 150 Page 1 of 8 Next

1: #257220. NIEMANN-PICK DISEASE, TYPE C1; NPC1 GeneTests, Links
NIEMANN-PICK DISEASE, TYPE D, INCLUDED
Gene map locus 18q11-q12

2: #123400. CREUTZFELDT-JAKOB DISEASE; CJD GeneTests, Links
CREUTZFELDT-JAKOB DISEASE, SPORADIC, INCLUDED; sCJD, INCLUDED
Gene map locus 6p21.3, 20pter-p12

3: #104300. ALZHEIMER DISEASE; AD GeneTests, Links
ALZHEIMER DISEASE, FAMILIAL, 1, INCLUDED; AD1, INCLUDED
Gene map locus 17q23.1, 17q23, 17q11.2, 12p11.23-q13.12, 12p13.3-p12.3, 11q23.2-q24.2, 10q24, 10q24, 7q36, 7q36, 7q36, 6p21.3, 4p14, 21q21, 20p, 19p13.2

4: #607822. ALZHEIMER DISEASE 3 GeneTests, Links
ALZHEIMER DISEASE, FAMILIAL, 3, WITH SPASTIC PARAPARESIS AND UNUSUAL PLAQUES, INCLUDED
Gene map locus 14q24.3

5: +107741. APOLIPOPROTEIN E; APOE MGI, GeneTests, Links
APOLIPOPROTEIN E, DEFICIENCY OR DEFECT OF, INCLUDED
Gene map locus 19q13.2

6: #127750. DEMENTIA, LEWY BODY; DLB Links
DIFFUSE LEWY BODY DISEASE WITH GAZE PALSY, INCLUDED
Gene map locus 5q35, 4q21

7: #168600. PARKINSON DISEASE; PD GeneTests, Links
Gene map locus 17q21.1, 11p15.5, 9q34, 8p22-p21.3, 6q27, 1p32, 4q22, 4p14, 2q22-q23, 2p13, Xq21-q25, 18p11.31-p11.2

8: +104760. AMYLOID BETA A4 PRECURSOR PROTEIN; APP MGI, GeneTests, Links
DEMENTIA, EARLY-ONSET PROGRESSIVE, AUTOSOMAL RECESSIVE, INCLUDED
Gene map locus 21q21

9: *176640. PRION PROTEIN; PRNP MGI, GeneTests, Links
Gene map locus 20pter-p12

培训班

MIM ID #257220

GeneTests, Links

NIEMANN-PICK DISEASE, TYPE C1; NPC1

Alternative titles; symbols

NIEMANN-PICK DISEASE, TYPE C; NPC
NIEMANN-PICK DISEASE WITH CHOLESTEROL ESTERIFICATION BLOCK
NIEMANN-PICK DISEASE, SUBACUTE JUVENILE FORM
NIEMANN-PICK DISEASE, CHRONIC NEURONOPATHIC FORM
NIEMANN-PICK DISEASE WITHOUT SPHINGOMYELINASE DEFICIENCY
NEUROVISCERAL STORAGE DISEASE WITH VERTICAL SUPRANUCLEAR OPHTHALMOPLEGIA

Other entities represented by this entry

NIEMANN-PICK DISEASE, TYPE D, INCLUDED
NIEMANN-PICK DISEASE, NOVA SCOTIAN TYPE, INCLUDED

Gene map locus: 18q11-q12

Clinical Synopsis

Text

Back to Top

A number sign (#) is used with this entry because Niemann-Pick disease type C1 and Niemann-Pick disease type D, also known as the Nova Scotian type, are caused by mutation in the NPC1 gene (607623).

Description

Back to Top

Niemann-Pick type C (NPC) disease is an autosomal recessive lipid storage disorder characterized by progressive neurodegeneration. Approximately 95% of cases are caused by mutations in the NPC1 gene, referred to as type C1; 5% are caused by mutations in the NPC2 gene (601015), referred to as type C2 (607625). The clinical manifestations of types C1 and C2 are similar because the respective genes are both involved in egress of lipids, particularly cholesterol, from late endosomes or lysosomes (Vance, 2006).

Historically, Crocker (1961) delineated 4 types of Niemann-Pick disease: the classic infantile form (type A; 257200), the visceral form (type B; 607616), the subacute or juvenile form (type C), and the Nova Scotian variant (type D). Types C1 and D are indistinguishable except for the occurrence of type D in patients of Nova Scotian Acadian ancestry. Since then, types E and F have also been described (see 607616), and phenotypic variation within each group has also been described.

Clinical Features

Back to Top

Niemann-Pick disease type C has a highly variable clinical phenotype. Patients with the 'classic' childhood onset type C usually appear normal for 1 or 2 years with symptoms appearing between 2 and 4 years. They gradually develop neurologic abnormalities which are initially manifested by ataxia, grand mal seizures, and loss of previously learned speech. Spasticity is striking and seizures, particularly myoclonic jerks, are common. Other features include dystonia, vertical supranuclear gaze palsy, dementia, and psychiatric manifestations. In general, hepatosplenomegaly is less striking than in types A and B, although it can be lethal in some. Cholestatic jaundice occurs in some patients. Foamy Niemann-Pick cells and 'sea-blue' histiocytes with distinctive histochemical and ultrastructural appearances are found in the bone marrow. In the

Table of Contents

Links

Selected Gene Related Links
Entrez Gene
Nomenclature
RefSeq
GenBank
Protein
UniGene
Other NCBI Links
BioSystems
Books
Free in PMC
GEO Profiles
Gene
Gene Genotype
GeneTests
GeneView in dbSNP
HomoloGene
Map Viewer
OMIA
PubMed (calculated)
PubMed (cited)
UniGene
Related Entries
EST
Nucleotide
Protein

NCBI

OMIM
Online Mendelian Inheritance in Man

Johns Hopkins University

My NCBI [Sign In] [Register]

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | PMC | OMIM

Search OMIM for [ ] Go Clear

Limits | Preview/Index | History | Clipboard | Details

Entrez

OMIM
Search OMIM
Search Gene Map
Search Morbid Map

Help
OMIM Help
How to Link

FAQ
Numbering System Symbols
How to Print
Citing OMIM
Download

OMIM Facts
Statistics
Update Log
Restrictions on Use

Allied Resources

- Enter one or more search terms.
- Use **Limits** to restrict your search by search field, chromosome, and other criteria.
- Use **Index** to browse terms found in OMIM records.
- Use **History** to retrieve records from previous searches, or to combine searches.

OMIM® - Online Mendelian Inheritance in Man®

Welcome to OMIM®, Online Mendelian Inheritance in Man®. OMIM is a comprehensive, authoritative, and timely compendium of human genes and genetic phenotypes. The full-text, referenced overviews in OMIM contain information on all known mendelian disorders and over 12,000 genes. OMIM focuses on the relationship between phenotype and genotype. It is updated daily, and the entries contain copious links to other

This da Mendelian
Inherita was created
in 198 s. It was
made g National
Center

OMIM icine, under
the dir

NCBI

OMIM
Online Mendelian Inheritance in Man

Johns Hopkins University

PubMed | Nucleotide | Protein | Genome | Structure | PopSet | Taxonomy | OMIM

The OMIM Gene map presents the cytogenetic map location of disease genes and other expressed genes described in OMIM. See the OMIM Morbid Map for a list of disease genes organized by disease. For more refined maps of genes and DNA segments click on the **Location** to invoke NCBI Entrez Map Viewer.

Search for: [ ] Find

- Enter gene symbol, chromosomal location, or disorder keyword to search for, e.g. "CYP1", "5", "1pter", "Xq", or "alzheimer".
- You must capitalize X and Y to search for those chromosomes.

The OMIM Gene map presents the cytogenetic map location of disease genes and other expressed genes described in OMIM. See the OMIM Morbid Map for a list of disease genes organized by disease. For more refined maps of genes and DNA segments click on the **Location** to invoke NCBI Entrez Map Viewer.

Search for: CYP1 Find Find Next (from the current location)

- Enter gene symbol, chromosomal location, or disorder keyword to search for, e.g. "CYP1", "5", "1pter", "Xq", or "alzheimer".
- You must capitalize X and Y to search for those chromosomes.

2p22-p21, CYP1B1 to 2p21, EHD3

<<Move Up Move Down>>

Location	Symbol	Title	MIM #	Disorder	Comments	Method	Mouse
2p22-p21	CYP1B1, GLC3A	Cytochrome P450, subfamily I, dioxin-inducible, polypeptide 1	601771	Glaucoma 3A, primary congenital, 231300 (3); Peters anomaly, 604229(3); Glaucoma, early-onset, digenic (3); Glaucoma, primary open angle, adult-onset, 137760 (3); Glaucoma, primary open angle, juvenile-onset, 137750(3)		REa, REn, A, Fd	
2p22-p21	EML4, ROPP120	Echinoderm microtubule associated protein like-4	607442			R, A	
2p22-p21	MEMO1, MEMO, C2orf4	Mediator of cell motility 1	611786			REc	
2p22-p21	MSH2, COCA1, FCC1, HNPCC1	mutS, E. coli, homolog of, 2	609309	Colorectal cancer, hereditary nonpolyposis, type 1, 120435 (3);Muir-Torre syndrome, 158320 (3); Mismatch repair cancer syndrome, 276300 (3)		Fd, REa, Ch	
2p22-p21	PUM2, KIAA0235	Pumilio, Drosophila, homolog of, 2	607205			REa, REc	
2p22-p21	SFRS7	Splicing factor, arginine/serine-rich 7, 35kD	600572			A	
2p22-p21	SLC30A6, ZNT6	Solute carrier family 30 (zinc transporter), member 6	611148			REc, H	17(Slc30a6)
2p22-p21	SOS1, GINGF, GF1, HGF, NS4	Son of sevenless, Drosophila, homolog of, 1	182530	Fibromatosis, gingival, 135300 (3); Noonan syndrome 4, 610733 (3);Noonan-like/multiple giant cell lesion syndrome, 163955 (3)		A, Fd, REc	17(Sos1)
2p22-p21	SPAST, SPG4	Spastin	604277	Spastic paraplegia-4, 182601 (3)		Fd	
2p22-p21	THUMPD2, C2orf8	THUMP domain-containing 2	611751			A	

OMIM®

Online Mendelian Inheritance in Man®

An Online Catalog of Human Genes and Genetic Disorders

Updated 5 August 2011

Search Sample Searches

Advanced Search: OMIM, Clinical Synopses, OMIM Gene Map

McKusick-Nathans Institute of Genetic Medicine

JOHNS HOPKINS MEDICINE

OMIM数据的下载

- ftp://ftp.ncbi.nih.gov/repository/OMIM/

  其中包含全部的OMIM的文件（omim.txt.Z），OMIM中的基因文件（genemap）及其解释文件（genemap.key），以及疾病信息（morbidmap）

- OMIM还提供genemap和morbidmap的网络查询形式

17,20-lyase deficiency, isolated, 202110 (3)	CYP17A1, CYP17, P450C17	609300	10q24.3
17-alpha-hydroxylase/17,20-lyase deficiency, 202110 (3)	CYP17A1, CYP17, P450C17	609300	10q24.3
1p36 deletion syndrome (2)	SKI	164780	1p36.3
2-methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency, 300438 (3)	HSD17B10, HADH2, ERAB, MRXS10, MRX17, MRX31, DUPXp11.22	300256	Xp11.2
2-methylbutyrylglycinuria, 610006 (3)	ACADSB, SBCAD	600301	10q25-q26
3-M syndrome, 273750 (3)	CUL7	609577	6p21.1
3-Methylcrotonyl-CoA carboxylase 1 deficiency, 210200 (3)	MCCC1, MCCA	609010	3q25-q27
3-Methylcrotonyl-CoA carboxylase 2 deficiency, 210210 (3)	MCCC2, MCCB	609014	5q12-q13
3-beta-hydroxysteroid dehydrogenase, type II, deficiency (3)	HSD3B2	201810	1p13.1
3-hydroxyacyl-CoA dehydrogenase deficiency, 231530 (3)	HADHSC, SCHAD, HHF4	601609	4q22-q26
3-hydroxyisobutryl-CoA hydrolase deficiency, 250620 (3)	HIBCH	610690	2q32.2
3-methylglutaconic aciduria, type I, 250950 (3)	AUH	600529	Chr.9
3-methylglutaconic aciduria, type III, 258501 (3)	OPA3, MGA3	606580	19q13.2-q13.3
3-methylglutaconic aciduria, type V, 610198 (3)	DNAJC19, TIM14	608977	3q26.3
3q21q26 syndrome (1)	EVI1	165215	3q26
5-fluorouracil toxicity, 274270 (3)	DPYD, DPD	612779	1p22
6-mercaptopurine sensitivity, 610460 (3)	TPMT	187680	6p22.3
ABCD syndrome, 600501 (3)	EDNRB, HSCR2, ABCDS	131244	13q22
ACAD9 deficiency, 611126 (3)	ACAD9	611103	3q26
ACAT2 deficiency (1) (?)	ACAT2	100678	6q25.3-q26
ACTH deficiency, 201400 (2)	CRH	122560	8q13
ADULT syndrome, 103285 (3)	TP63, TP73L, KET, EEC3, SHFM4, LMS, RHS, OFC8	603273	3q27
AGAT deficiency, 612718 (3)	GATM, AGAT	602360	15q15.3
AICA-ribosiduria due to ATIC deficiency, 608688 (3)	ATIC, PURH, AICAR	601731	2q35
ARC syndrome, 208085 (3)	VPS33B	608552	15q26.1
Aarskog-Scott syndrome, 305400 (3)	FGD1, FGDY, AAS	300546	Xp11.21
Abdominal obesity-metabolic syndrome (2)	AOMS1, SYNX	605552	3q27
Abdominal obesity-metabolic syndrome (2)	AOMS2	605572	17p12
Abetalipoproteinemia, 200100 (3)	MTP	157147	4q22-q24
Acampomelic campomelic dysplasia, 114290 (3)	SOX9, CMD1, SRA1	608160	17q24.3-q25.1

(1) the disorder was positioned by mapping of the wildtype gene;
(2) the disease phenotype itself was mapped;
(3) the molecular basis of the disorder is known;
(4) the disorder is a chromosome deletion or duplication syndrome.

遗传关联数据库（GAD）

某一疾病相关的基因

- GAD (Genetic Association Database)
- GAD的发展史

  由美国国立卫生研究院（National Institutes of Health，NIH）开发和维护
- 可以通过http://geneticassociationdb.nih.gov/ 访问该数据库

选择“Browser All”链接可以得到如下结果

Genetic Association Database

Batch Search Results

Gene	Total	AGE	CAN	CARD	CHEM	DEV	HEM	IMM	INF	MET	MITO	NEUR	NV	PHARM	PSY	REN	REP	VIS	Other	Unknown
HESX1	3	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	2	-
HAVCR1	5	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-
HMHA1	2	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-
GYS2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
H6PD	3	-	-	-	-	-	-	-	-	2	-	1	-	-	-	-	-	-	-	-
GUCA1A	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
GSTT2	3	-	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GSTO2	10	-	5	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-
GSTM3	40	-	29	-	1	-	-	2	-	-	-	2	-	1	-	-	-	1	3	1
GRIN2D	2	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
GRIN2B	25	-	-	-	6	-	-	-	-	-	-	4	-	-	14	-	-	-	-	1
GRIN2C	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
GRIK3	8	-	-	-	1	-	-	-	-	-	-	-	-	-	6	-	-	-	-	1
GRHPR	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
GPX1	38	-	22	6	-	-	-	1	-	2	-	-	-	1	1	2	-	1	1	1
GNB1	2	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-
GJB2	51	-	-	-	-	-	-	-	1	-	-	-	2	-	-	-	-	-	48	-
GFRA3	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
GDAP1	4	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	2	-
GCGR	12	-	-	1	-	-	-	-	-	8	-	-	-	-	-	-	-	-	3	-
GCG	4	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-

从数据库获取糖尿病相关基因

Genetic Association Database

Query Disease View

Type in search conditions for each field. You can use >=, >, <=, <, 34.2..45.0, any text, in the search field. Blank field will be default to return all records. Press submit button when done.

可限定，GAD是个关联分析数据库，所有有个关联分析的p值，OMIM是个实验得到的数据库，GAD是通过系统统计得到的数据库

Broad Phenotype (Disease): diabetes

Disease Class: Metabolic

Narrow Phenotype:

Molecular Phenotype:

P-value: 0. 01

Chr:

Ch-Band:

DNA start position:

DNA end position:

Gene Symbol (Official):

Reference:

PubMed Link:

PubMed+Gene+Disease:

OMIM record ID

Association? Y/N:

Search Clear

Cardiovascular
Chemical Dependency
Developmental
Hematological
Immune
Infection
Metabolic
Mitochondrial
Neurological
Normal Variation
Other
Pharmacogenomic
Psychiatric
Renal
Reproduction
Vision
Unknown

	Assoc? YorN	Gene Symbol	Disease Class	Broad Phenotype (Disease)	MeSH Disease Terms
view	Y	BCHE	METABOLIC	diabetes, type 2	Diabetes Mellitus, Type 2|Genetic Predisposition to Disease
view	Y	LTA	METABOLIC	diabetes, type 2	Diabetes Mellitus, Type 2|Hypertriglyceridemia
view	Y	TCF7L2	METABOLIC	diabetes, type 2	Diabetes Mellitus, Type 2|
view	Y	IRS1	METABOLIC	diabetes, type 2	Diabetes Mellitus, Type 2|
view	Y	TCF7L2	METABOLIC	diabetes, type 2	Glucose Metabolism Disorders|Diabetes Mellitus, Type 2|Insulin Resistance|Disease Progression
view	Y	IL1B	METABOLIC	prevalent Type 2 Diabetes and	Diabetes Mellitus

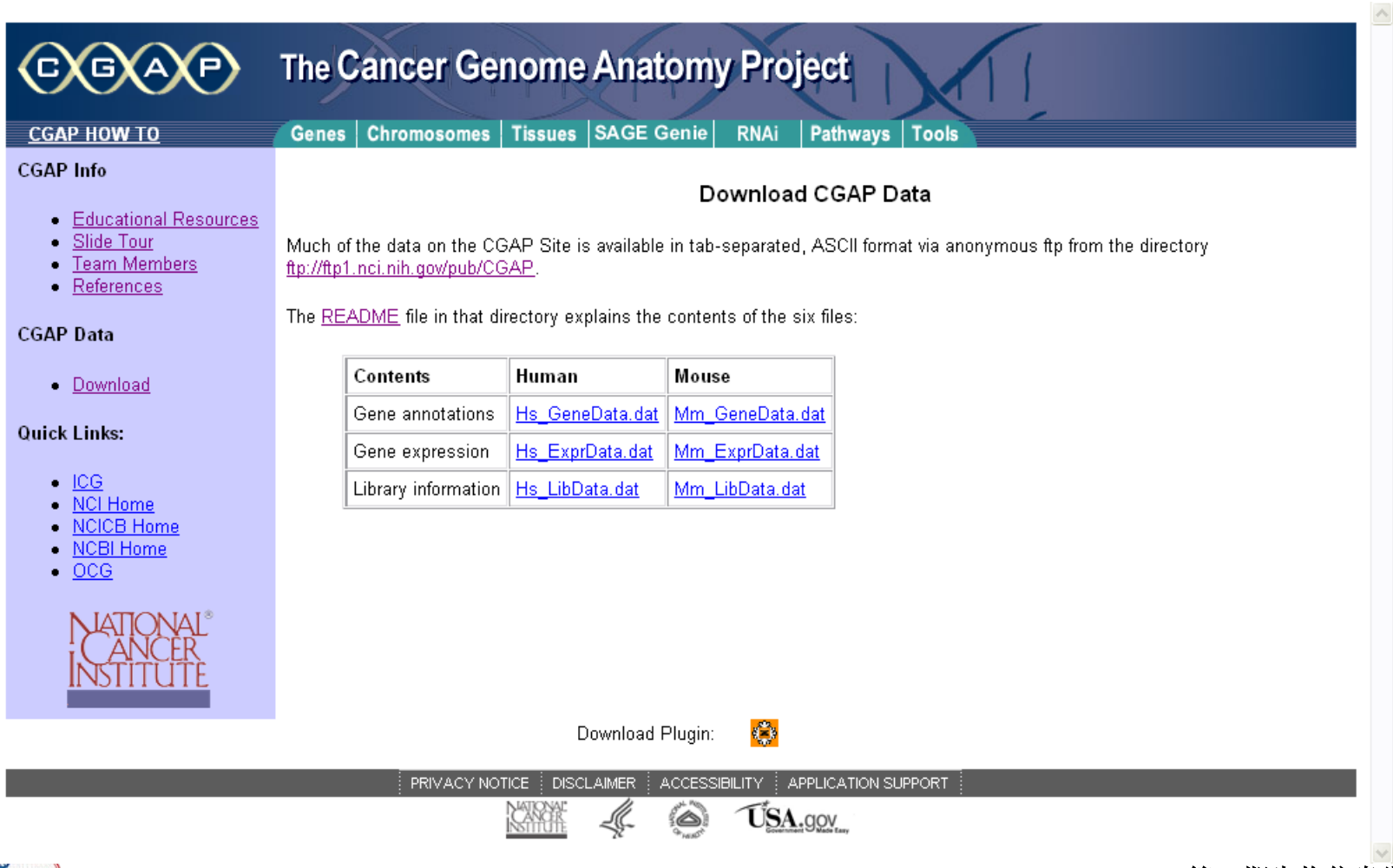
P Value	Chr	Ch-Band	DNA Start(bp)	DNA End(bp)	Reference	PUB MED	PUB GN DS	OMIM
p=0.00017	3	3q26.1-q26.2	165490693	165555253	Hashim Y 2001	PM	PM+ GAD	177400
p < 0.001	6	6p21.3	31540092	31542098	Vendrell J1995	PM	PM+ GAD	153440
0.001	10	10q25.3	114710008	114927434	P M Thorsby , et al. Scandinavian journal of clinical and laboratory investigation 2009 69(2):282-7	PM	PM+ GAD	602228
0.0001	2	2q36	227596033	227663506	Ana I Burguete-Garcia , et al. Metabolism 2010 59(1):38-45	PM	PM+ GAD	147545
0.001	10	10q25.3	114710008	114927434	Florez, J.C. et al. 2006	PM	PM+ GAD	602228
p=0.037, P-gluc: p less 0.0001(single rs1143634)	2	2q14	113587336	113594356	Kari Luotola , et al. The Journal of clinical endocrinology and metabolism 2009 94(11):4575-83	PM	PM+ GAD	147720

癌症基因数据库（CGAP）

通过芯片差异分析得到的，有

- Cancer Genome Anatomy Project，癌基因组解剖计划是一项由美国癌症研究所（National Cancer Institute，NCI）于1996年发起并建立和主持的交叉学科计划
- 用户可以通过http://cgap.nci.nih.gov/进行访问
- CGAP的创建目的及总体目标

下载页面图所示，其中包含了人和小鼠两个物种的基因注释、基因表达以及相关的一些文库中的数据。



WHO规范的疾病分类标准（ICD）

- 国际疾病分类，简称ICD（International Classification of Diseases），是目前国际上共同使用的统一的疾病分类方法。
- ICD的目的是对不同国家或地区在不同时间收集到的死亡和疾病数据进行系统地记录、分析、解释和比较，其中包括对各人群组一般健康状况的分析，疾病发病和患病的监测以及其有关的其他健康问题。

ICD-10编码查询 (国际疾病分类编码）

在WHO网站上，只是个数据表，分为22类

Chapter	Blocks	Title
I	A00-B99	Certain infectious and parasitic diseases
II	C00-D48	Neoplasms
III	D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	E00-E90	Endocrine, nutritional and metabolic diseases
V	F00-F99	Mental and behavioural disorders
VI	G00-G99	Diseases of the nervous system
VII	H00-H59	Diseases of the eye and adnexa
VIII	H60-H95	Diseases of the ear and mastoid process
IX	I00-I99	Diseases of the circulatory system
X	J00-J99	Diseases of the respiratory system
XI	K00-K93	Diseases of the digestive system
XII	L00-L99	Diseases of the skin and subcutaneous tissue
XIII	M00-M99	Diseases of the musculoskeletal system and connective tissue
XIV	N00-N99	Diseases of the genitourinary system
XV	O00-O99	Pregnancy, childbirth and the puerperium
XVI	P00-P96	Certain conditions originating in the perinatal period
XVII	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities
XVIII	R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
XIX	S00-T98	Injury, poisoning and certain other consequences of external causes
XX	V01-Y98	External causes of morbidity and mortality

Disease Ontology

- 2003年在Northwestern大学启动的Nugene计划的一部分
- 提供一个与人类疾病相关的整合的生物医学数据集的开源Ontology体系
- 促进各种疾病及相关健康状况向特定医学代码的映射
- 期望构建成一个具有正确的Ontology体系结构并且在语义上可计算的结构形式

DO的构成

DO是将不同数据库通过疾病概念整合到一起的开源的疾病体系：

Medical Subject Headings（MeSH）

Universal Medical Language System（UMLS）

International Classification of Disease（ICD）

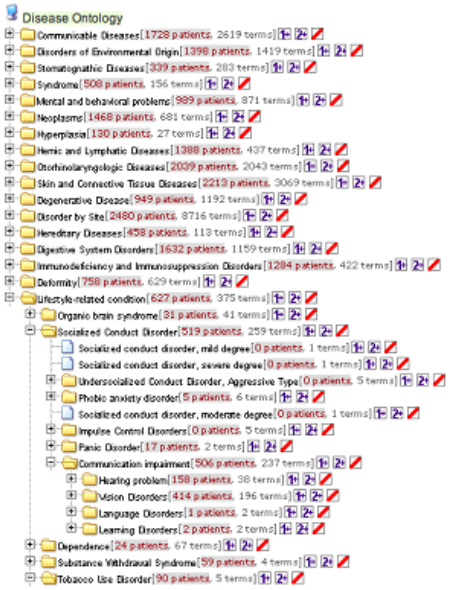
Systematized Nomenclature of Human Veterinary Medicine—Clinical Term

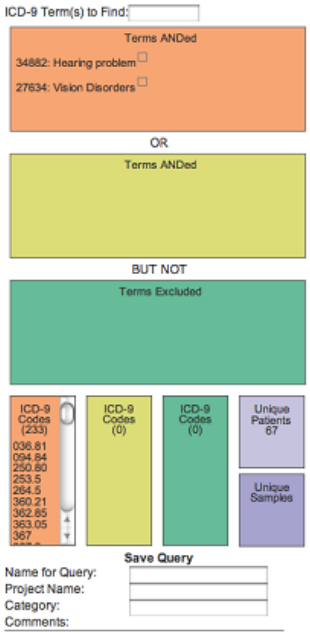
下载下来才可以可视化

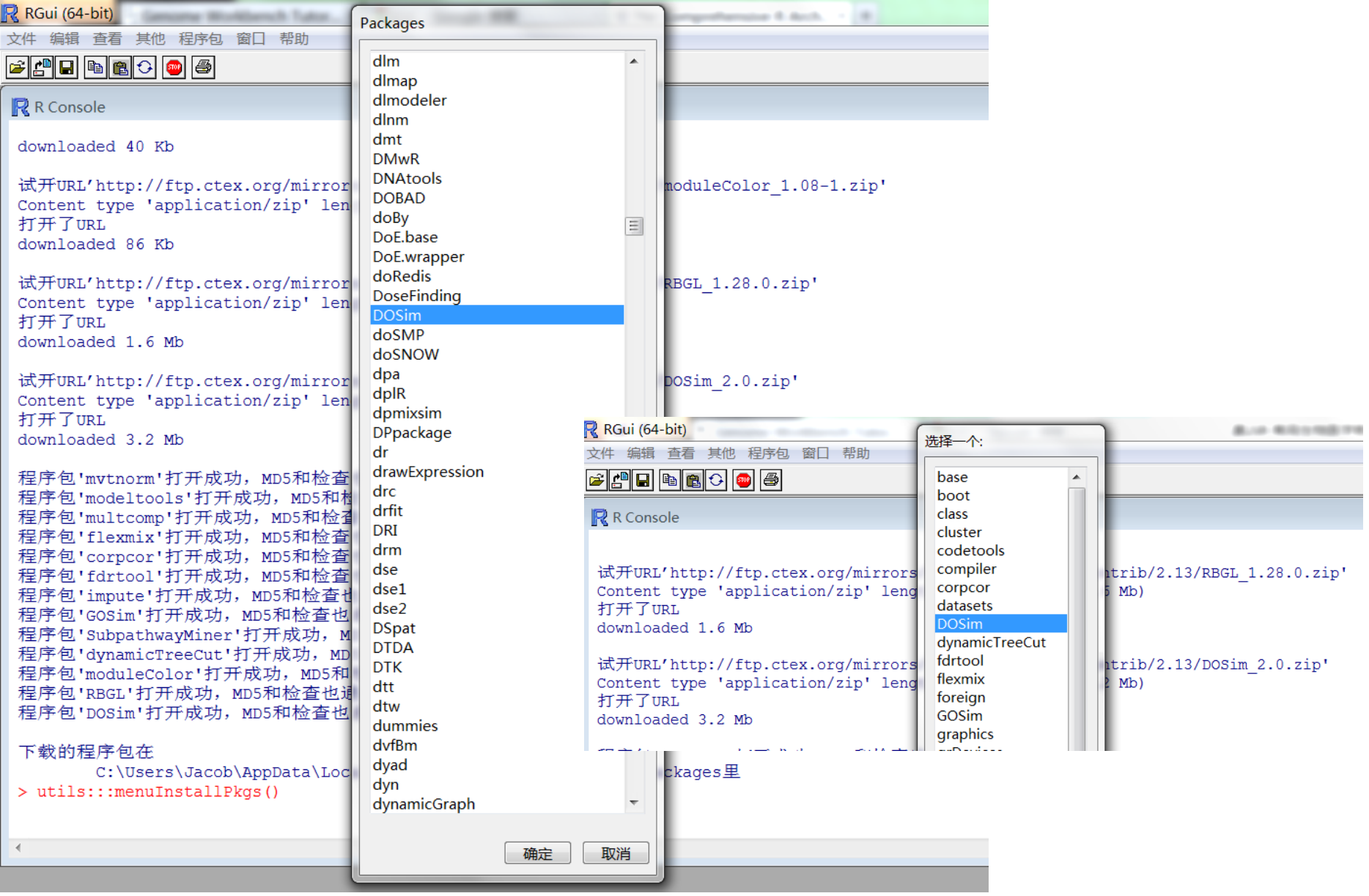
DO Mappings

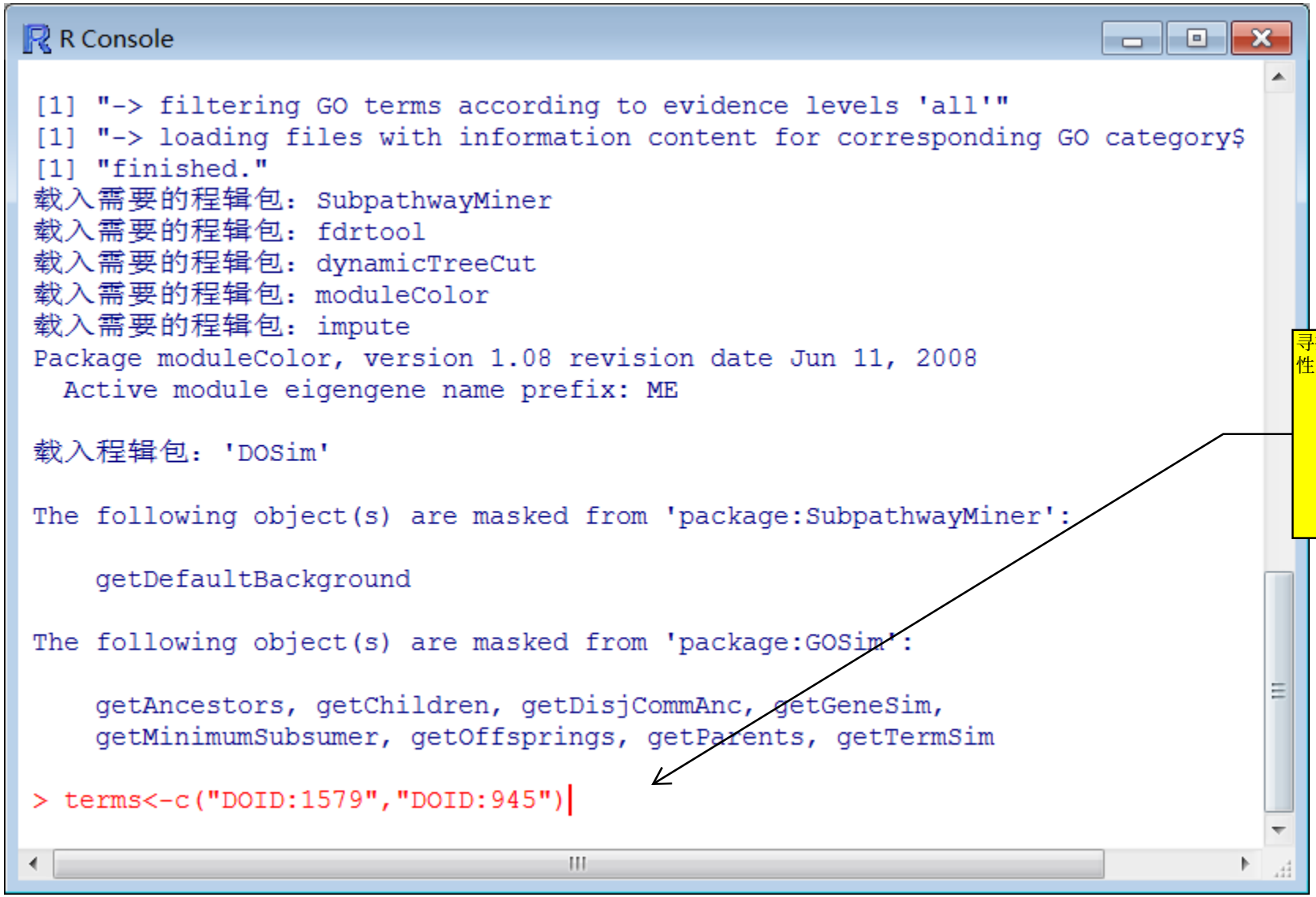
External Reference	Unique xref:DOID Mappings	Unique xrefs
ICD-9	186278	10109
UMLS_SNOM DCT_2005_01_31_AUI	38912	38912
UMLS_NCI2004_11_17_AUI	24049	24049
UMLS_MSH2005_2005_01_17_AUI	21377	21377
UMLS_CUI	17023	17023
UMLS_ST	14674	14674
SNOMEDCT_2005_01_31	13116	13116
UMLS_ICD-9	10048	10048
NCI2004_11_17	6991	6991
UMLS_MTHICD-9_2005_AUI	3611	3611
MSH2005_2005_01_17	3502	3502
UMLS_CSP2004_AUI	2269	2269

Example Application - DO Browser






RGui (64-bit)
文件 编辑 查看 其他 程序包 窗口 帮助
R Console
downloaded 40 Kb
试开URL'http://ftp.ctex.org/mirror
moduleColor_1.08-1.zip'
Content type 'application/zip' len
打开了URL
downloaded 86 Kb
试开URL'http://ftp.ctex.org/mirror
RBGL_1.28.0.zip'
Content type 'application/zip' len
打开了URL
downloaded 1.6 Mb
试开URL'http://ftp.ctex.org/mirror
DOSim_2.0.zip'
Content type 'application/zip' len
打开了URL
downloaded 3.2 Mb
程序包'mvtnorm'打开成功，MD5和检查
程序包'modeltools'打开成功，MD5和
程序包'multcomp'打开成功，MD5和检查
程序包'flexmix'打开成功，MD5和检查
程序包'corpcor'打开成功，MD5和检查
程序包'fdrtool'打开成功，MD5和检查
程序包'impute'打开成功，MD5和检查也
程序包'GOSim'打开成功，MD5和检查也
程序包'SubpathwayMiner'打开成功，M
程序包'dynamicTreeCut'打开成功，MD
程序包'moduleColor'打开成功，MD5和
程序包'RBGL'打开成功，MD5和检查也通
程序包'DOSim'打开成功，MD5和检查也
下载的程序包在
C:\Users\Jacob\AppData\Loc
ckages里
> utils:::menuInstallPkgs()
Packages
dlm
dlmap
dlmodeler
dlnm
dmt
DMwR
DNAtools
DOBAD
doBy
DoE.base
DoE.wrapper
doRedis
DoseFinding
DOSim
doSMP
doSNOW
dpa
dplR
dpmixsim
DPpackage
dr
drawExpression
drc
drfit
DRI
drm
dse
dse1
dse2
DSpat
DTDA
DTK
dtt
dtw
dummies
dvfBm
dyad
dyn
dynamicGraph
确定
取消
RGui (64-bit)
文件 编辑 查看 其他 程序包 窗口 帮助
R Console
试开URL'http://ftp.ctex.org/mirrors
trib/2.13/RBGL_1.28.0.zip'
Content type 'application/zip' leng
Mb)
打开了URL
downloaded 1.6 Mb
试开URL'http://ftp.ctex.org/mirrors
trib/2.13/DOSim_2.0.zip'
Content type 'application/zip' leng
Mb)
打开了URL
downloaded 3.2 Mb
选择一个:
base
boot
class
cluster
codetools
compiler
corpcor
datasets
DOSim
dynamicTreeCut
fdrtool
flexmix
foreign
GOSim
graphics


R Console
[1] "-> filtering GO terms according to evidence levels 'all'"
[1] "-> loading files with information content for corresponding GO category$
[1] "finished."
载入需要的程辑包: SubpathwayMiner
载入需要的程辑包: fdrtool
载入需要的程辑包: dynamicTreeCut
载入需要的程辑包: moduleColor
载入需要的程辑包: impute
Package moduleColor, version 1.08 revision date Jun 11, 2008
Active module eigengene name prefix: ME
载入程辑包: 'DOSim'
The following object(s) are masked from 'package:SubpathwayMiner':
getDefaultBackground
The following object(s) are masked from 'package:GOSim':
getAncestors, getChildren, getDisjCommAnc, getGeneSim,
getMinimumSubsumer, getOffsprings, getParents, getTermSim
> terms<-c("DOID:1579","DOID:945")
寻找两个疾病
性，共享的基

R Console

```
> print(res)
$`DOID:1579`
[1] "respiratory system disease"

$`DOID:1578`
[1] "pulmonary systemic sclerosis"

> tsim<-getTermSim(terms)
[1] "Terms: DOID:1579 , DOID:1579 ( method: relevance )"
[1] "Terms: DOID:1578 , DOID:1578 ( method: relevance )"
[1] "Terms: DOID:1578 , DOID:1579 ( method: relevance )"
> print(tsim)
          DOID:1579 DOID:1578
DOID:1579 0.7210163         0
DOID:1578 0.0000000         1
> getTermSim(terms, method = "Wang")
[1] "Terms: DOID:1579 , DOID:1579 ( method: Wang )"
[1] "Terms: DOID:1578 , DOID:157
[1] "Terms: DOID:1578 , DOID:157
          DOID:1579 DOID:1578
DOID:1579 1.0000000 0.3324382
DOID:1578 0.3324382 1.0000000
>
```

呼吸系统疾病

肺系统性硬化症

不是一是因为呼吸系统疾病很宽泛，共享的genelist是否可显示？

Currently the following methods for computing DO term similarities are implemented:

"Resnik" information content of minimum subsumer (ICms) [1]

"JiangConrath" $1 - \min(1, IC(term1) - 2ICms + IC(term2))$[2]

"Lin" $\frac{2ICms}{(IC(term1)+IC(term2))}$ [3]

"CoutoResnik" average information content of common disjunctive ancestors of `term1` and `term2` (ICshare) [4]

"CoutoJiangConrath" $1 - \min(1, IC(term1) - 2ICshare + IC(term2))$ [4]

"CoutoLin" $\frac{2ICshare}{(IC(term1)+IC(term2))}$ [4]

"relevance" sim_Lin * (1 - exp(-ICms)) [5]

默认测度

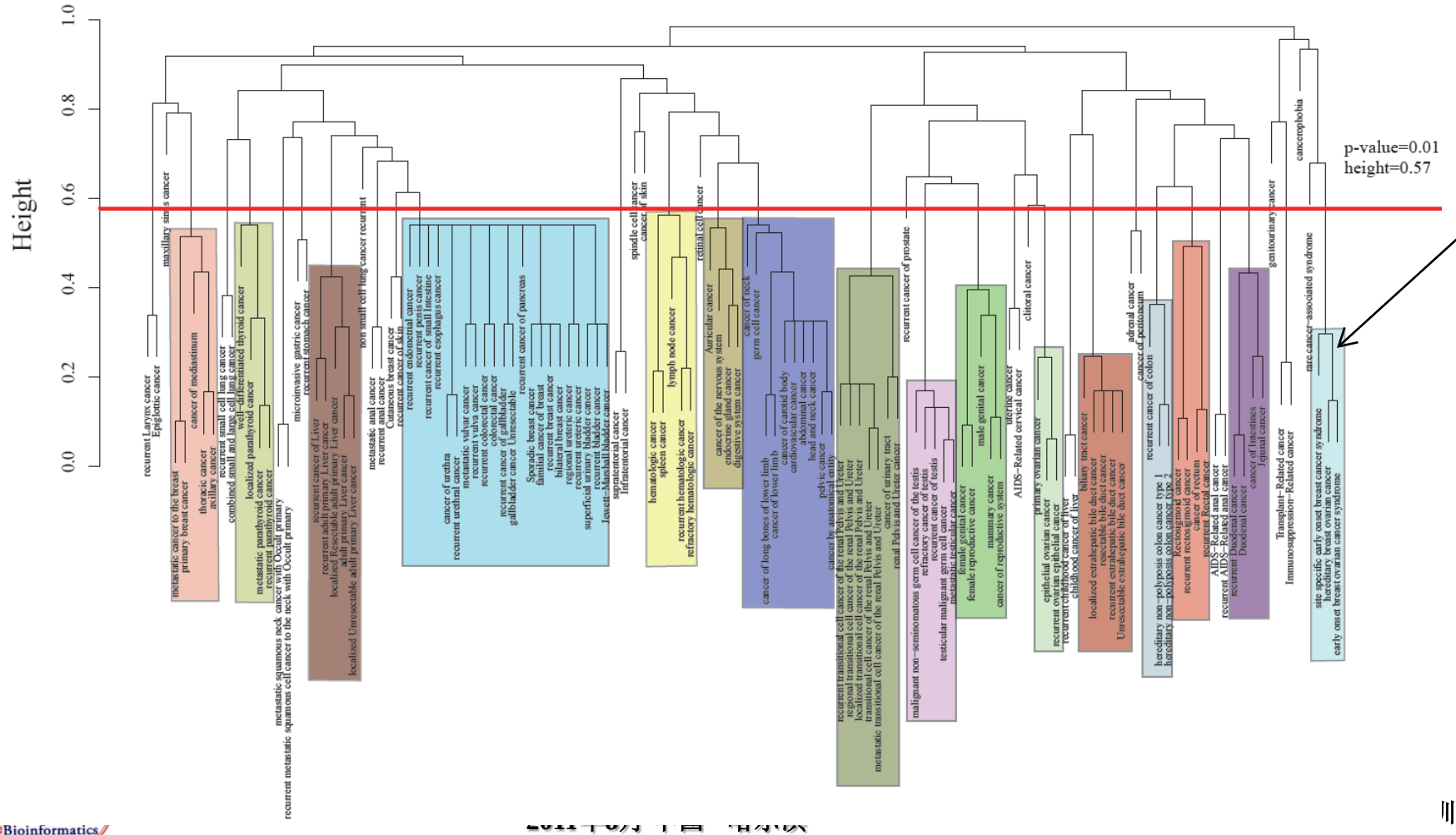
"GIC" summed information content of common ancestors divided by summed information content of all ancestors of term1 and term2 [7]

"simIC" sim_Lin * (1 - 1/(1 + ICms)) [7]

"Wang" $Sim(term1, term2) = \frac{\sum_{t \in T_{term1} \cap T_{term2}} (S_{term1}(t) + S_{term2}(t))}{SV(term1) + SV(term2)}$ [8]

Bioinformatics

DOSim: An R package for similarity between diseases based on Disease Ontology



Bioinformatics

III班

Thanks!

祝大家在浩瀚的生物医学数据资源中获得有用的信息，
在科学研究中取得更辉煌的成绩！