

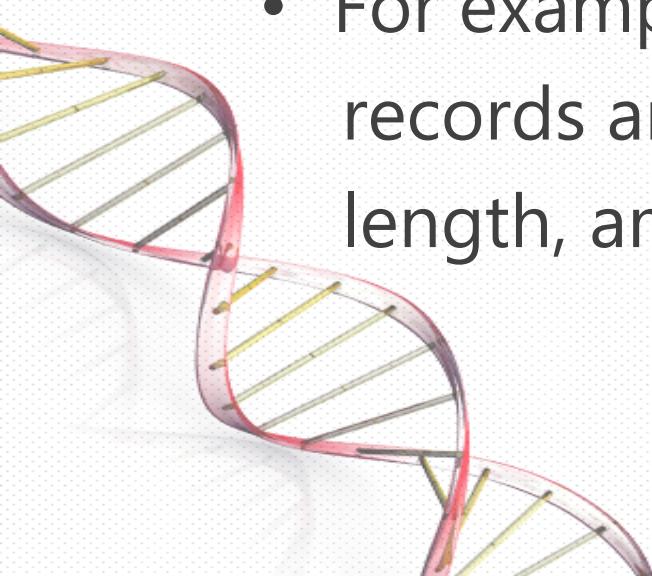


Biological Databases

Hammad Naveed
hammad.naveed@nu.edu.pk

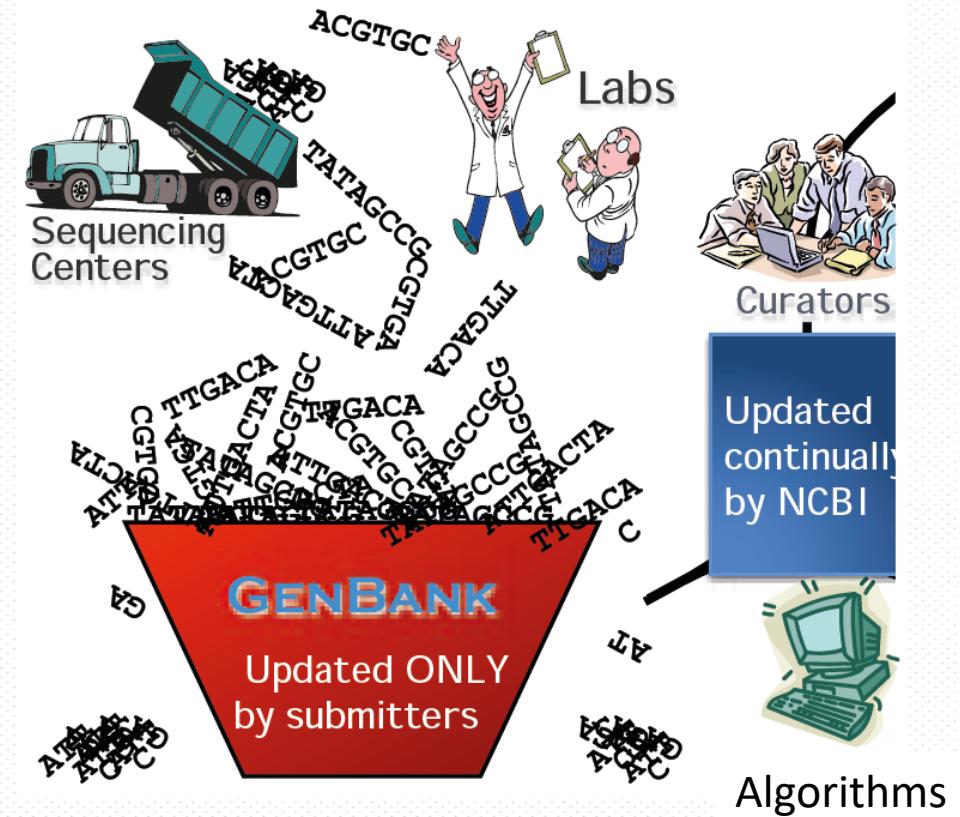
What is a database

- **Structured** collection of information.
- Consists of basic units called records or entries.
- Each record consists of fields, which hold pre-defined data related to the record.
- For example, a protein database would have protein entries as records and protein properties as fields (e.g., name of protein, length, amino-acid sequence)

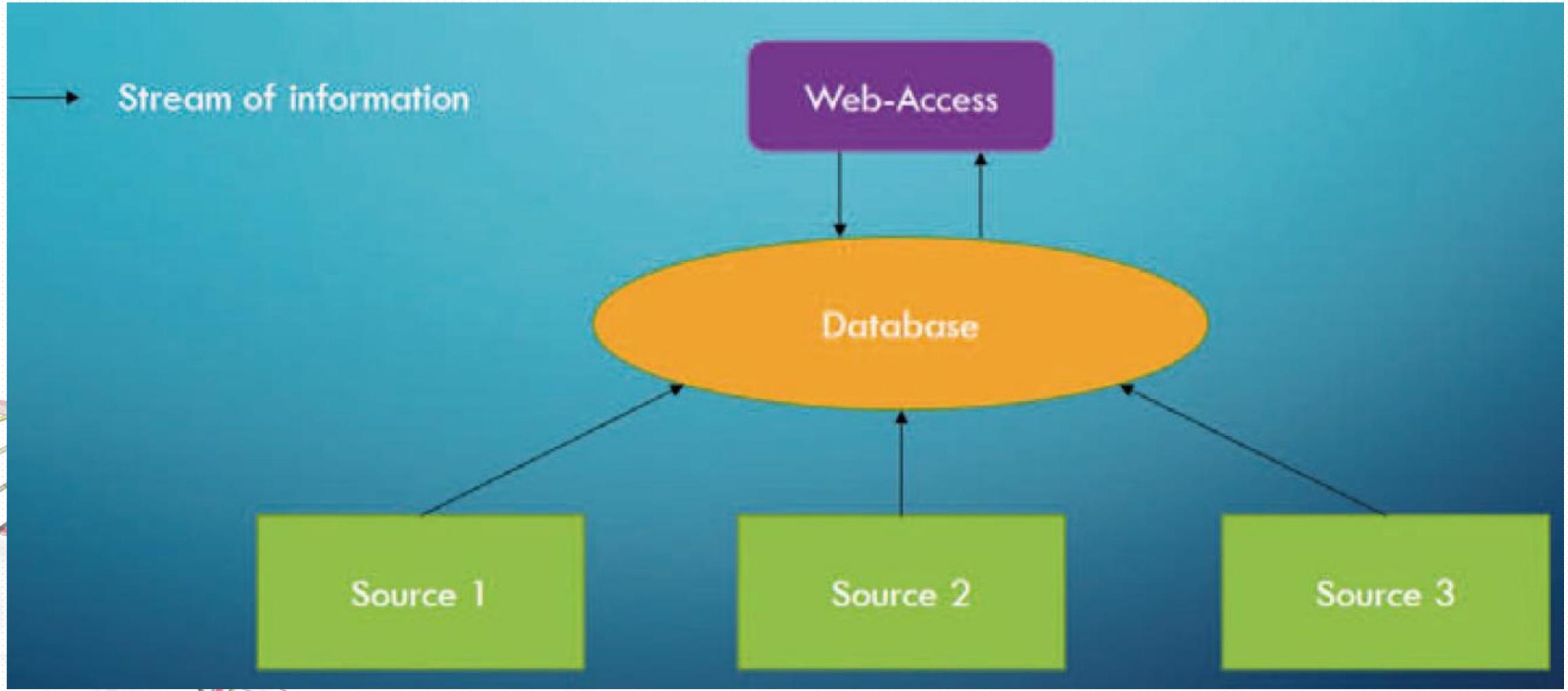


Types of databases

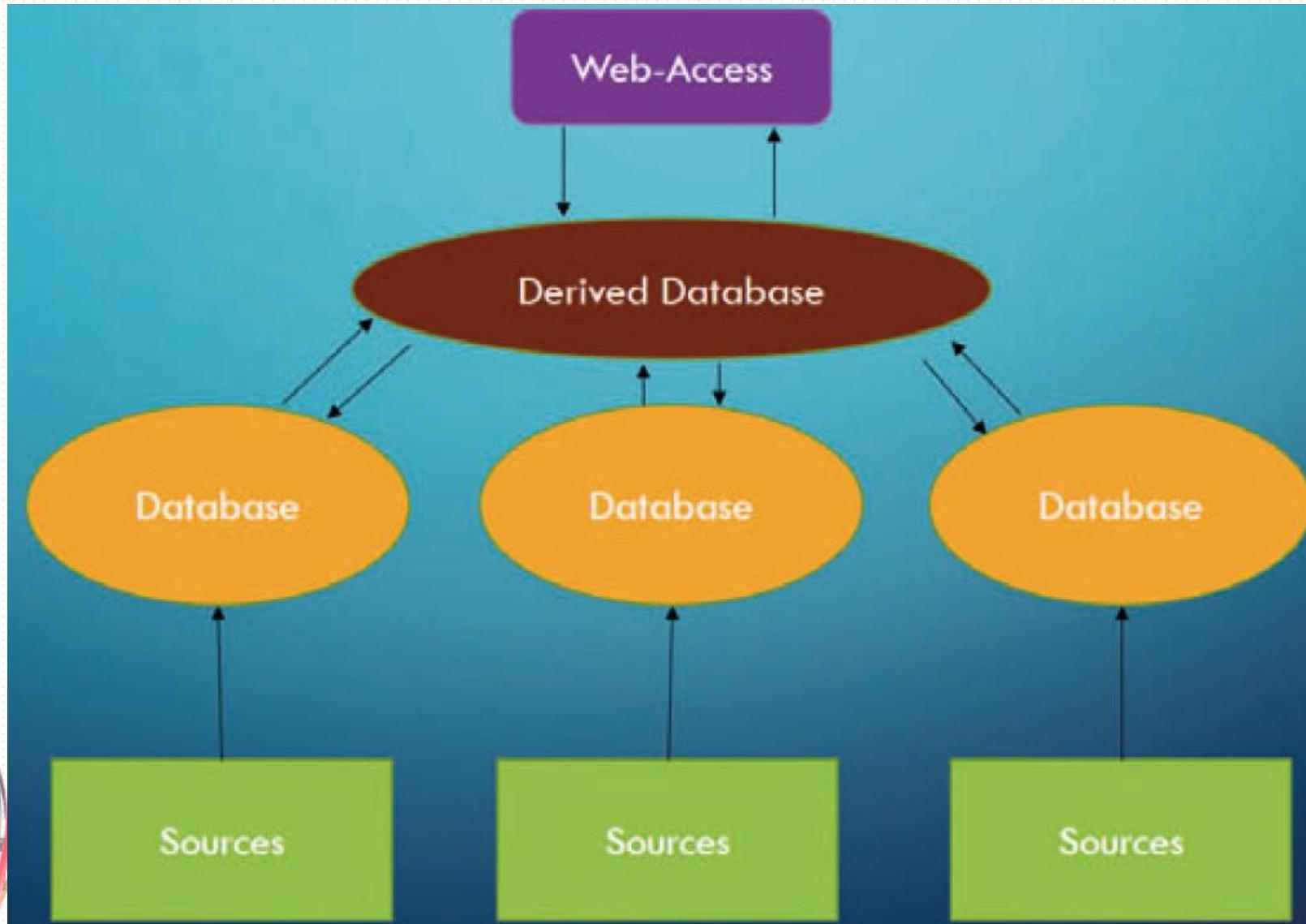
- Primary Databases
 - Original submissions by experimentalists
 - Content controlled by the **submitter**
 - Examples: GenBank, Trace, SRA, SNP, GEO
- Derivative Databases
 - Derived from primary data
 - Content controlled by **third party (NCBI)**
 - Examples: NCBI Protein, Refseq, TPA, RefSNP, GEO datasets, UniGene, Homologene, Structure, Conserved Domain



Primary databases



Secondary databases



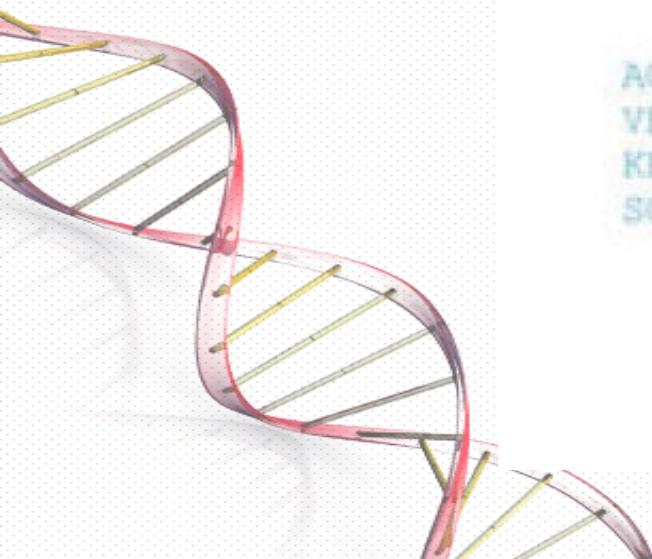
A flat-file database

(A)

NAME	TELEPHONE	ADDRESS
S. Claus	0203 450	The North Pole, Lapland
M. Mouse	0202 453	Disneyworld, Florida
A. Moonman	0104 459	Craterland, The Moon

(B) GenBank Flat-File Format

LOCUS SCU49845 5028 bp DNA
DEFINITION *Saccharomyces cerevisiae* TCPI-beta gene, partial cds, and Ax12p
(AXL2) and Rev7p (REV7) genes, complete cds.
ACCESSION U49845
VERSION U49845.1 GI:1293613
KEYWORDS
SOURCE *Saccharomyces cerevisiae* (baker's yeast)
ORGANISM *Saccharomyces cerevisiae*
Eukaryota; Fungi; Ascomycota; Saccharomycotina;
Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; *Saccharomyces*.



Relational databases

protab1			
Protein-code	Protein-name	Length	Species-origin
P1001	Hemoglobin	145	Bovine
P1002	Hemoglobin	136	Ovine
P1003	Eye Lens Protein	234	Human
.....			

protab2	
Protein-code	Protein-sequence
P1001	MDRTTHGFDLKLLSPRTVNQWLMLALFFGHS...
P1002	MDKTSHGFEIKLLTPKKLQQWLMIAlYFGHT...
P1003	SRTHEEEGKLMQWPPRPLYIALFTEPPYP...
.....	



Relational database

- A relational database consists of relations (tables) containing attributes (fields or columns). Each row in a table is known as a record or tuple.
- Information should be 'normalized' so that it is non-redundant this means that every row should be unique, although this ideal is not always observed.

First Name	Last Name	Institution	Department	Address
Omar	Farooq	NUCES	Computer Science	Islamabad
Hadiya	Ali	FAST	Electrical Engineering	Islamabad
Ahmed	Khan	NUCES	Dept of Computer Science	Isb
Ahmed	Khan	NUST	Dept of Management	Islamabad

Omar|Farooq|Computer Science|NUCES|Islamabad
 Hadiya|Ali|Electrical Engineering|FAST|Islamabad
 Ahmed|Khan|Dept of Computer Science|NUCES|Isb
 Ahmed|Khan|Dept of Management|NUST|Islamabad

Foreign Key

Primary Key

Table Contacts

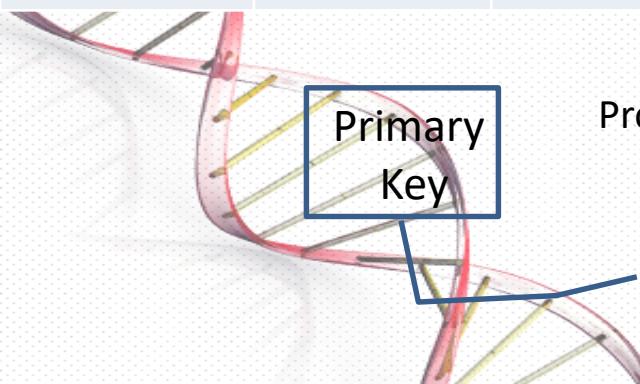


Table Professor

Prof_id	First_name	Last_name	Contact_id
1	Omar	Farooq	1
2	Hadiya	Ali	2
3	Ahmed	Khan	1
4	Ahmed	Khan	3

Contact_id	Institution	Department	Address
1	NUCES	Computer Science	Islamabad
2	FAST	Electrical Engineering	Islamabad
3	NUST	Management	Islamabad

Types of databases

DNA & RNA

genes, genomes & variation

Structures

Molecular & cellular structures

Ontologies

taxonomies & controlled vocabularies



Gene expression

RNA, protein & metabolite expression



Proteins

sequences, families & motifs



Systems

reactions, interactions & pathways



Chemical biology

chemogenomics & metabolomics



Literature

Scientific publications & patents

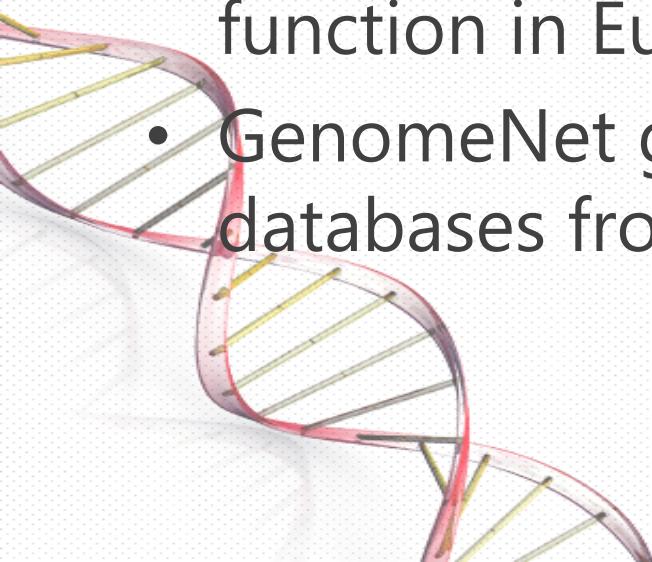


Other software

cross-domain tools & resources

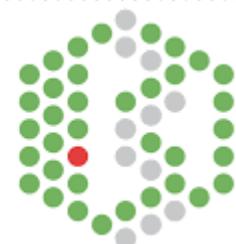
Database providers

- The National Center for Biotechnology Information (NCBI) offers data banks, databases and tools (USA)
- The European Bioinformatics Institute (EBI) does a similar function in Europe
 - GenomeNet gathers several databases from Japan



National
Center for
Biotechnology
Information

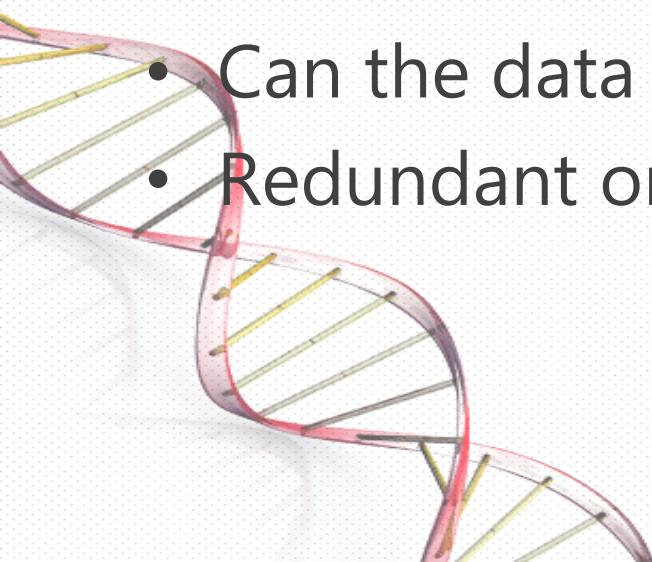
EMBL-EBI



DDBJ
DNA Data Bank of Japan

Data quality

- How are things entered
 - Step by step protocol
- What are the evidence?
 - Automatic validation
 - Manual curation
- How new is the data?
- Can the data be secret?
- Redundant or non-redundant?





National Center for
Biotechnology Information

[All Databases](#) ▾[Search](#)[NCBI Home](#)[Resource List \(A-Z\)](#)[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](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts
into NCBI databases



Download

Transfer NCBI data to your
computer



Learn

Find help documents, attend a
class or watch a tutorial



Develop

Use NCBI APIs and code
libraries to build applications



Analyze

Identify an NCBI tool for your
data analysis task



Research

Explore NCBI research and
collaborative projects



Popular Resources

[PubMed](#)[Bookshelf](#)[PubMed Central](#)[PubMed Health](#)[BLAST](#)[Nucleotide](#)[Genome](#)[SNP](#)[Gene](#)[Protein](#)[PubChem](#)

NCBI News & Blog

[Identical Protein Groups: Non-redundant access to protein records](#)

26 Jul 2017

[Have you ever searched the NCBI Protein database and been overwhelmed?](#)

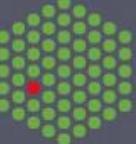
[ClinVar Allele-Based Summaries Now Available for FTP Download](#)

25 Jul 2017

[ClinVar: NCBI's archive of submitted](#)

European Bioinformatics Institute



EMBL-EBI 

The home for big data in biology

Our unique Search service helps you explore dozens of biological data resources.
[More about EBI Search ➤](#)

All Find a gene, protein or chemical

Example searches: [blast](#) [keratin](#) [bf1](#)

About this image: An entry from the Tara Oceans expedition in the prototype Image Data Repository, built by scientists at the University of Dundee, EMBL-EBI, the University of Bristol and the University of Cambridge. [Learn more ➤](#)

Find a tool for your data analysis.

Share your scientific data with the world.

Explore EMBL-EBI and our mission

The European Bioinformatics Institute (EMBL-EBI) shares [data from life science experiments](#), performs [basic research in computational biology](#) and offers an extensive [user training programme](#), supporting researchers in academia and industry. We are part of [EMBL](#), Europe's flagship laboratory for the life sciences.

[More about EMBL-EBI and our impact ➤](#)

Services

We provide freely available data and bioinformatics services to all facets of the scientific community [➤](#)

Research

We contribute to the advancement of biology through basic investigator-driven research [➤](#)

Training

We provide advanced bioinformatics training to scientists at all levels [➤](#)

Industry

We help disseminate cutting-edge technologies to industry [➤](#)

ELIXIR

We support, as an ELIXIR node, the coordination of biological data provision throughout Europe [➤](#)

GenomeNet

[English | Japanese]



Search for

Due to the network maintenance, GenomeNet service will not be available during the following time.
Fri, 11 Aug 2017 01:00-02:00 (JST)
Sorry for the inconveniences.

GenomeNet
[About GenomeNet](#)
[Release notes](#)
[Acknowledgments](#)

DBGET
[Overview](#)
[DB release info](#)

KEGG

varDB

Community DBs

Bioinformatics tools
[Other tools](#)

FTP

Feedback

GenomeNet Database Resources

DBGET: Integrated Database Retrieval System
[DBGET search](#)
[LinkDB search](#) SPARQL endpoint available

KEGG: Kyoto Encyclopedia of Genes and Genomes
[KEGG2 - Table of contents](#)
[KEGG PATHWAY](#) - Systems information: pathways
[KEGG BRITE](#) - Systems information: ontologies
[KEGG Organisms](#) - Organism-specific entry points
[KEGG GENES](#) - Genomic information
[KEGG LIGAND](#) - Chemical information
[KEGG MEDICUS](#) - Health information

KEGG MGENES: Metagenome gene catalogs
[Virus-Host DB](#): Hosts of sequenced viruses

Taxonomy: Organism classification

Reaction Ontology: Reaction classifications

varDB: Antigenic variation database

Community Databases

[CYORF](#) - Cyanobacteria annotation database
[BSORF](#) - Bacillus subtilis genome database
[EXPRESSION](#) - Gene expression profile database

BRITE Functional Hierarchies
[KEGG Pathway Maps](#)
[KEGG Organisms](#)

Genome Catalogs
[Complete \(KEGG GENES\)](#)
[Meta \(KEGG MGENES\)](#)
[FTP download](#)
[Viral \(Viruses\)](#)
[EST \(EGENES\)](#)

Ortholog Clusters (OC)



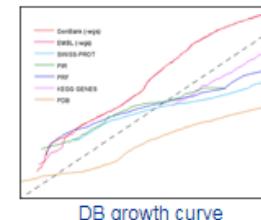
Database links

GenomeNet Bioinformatics Tools

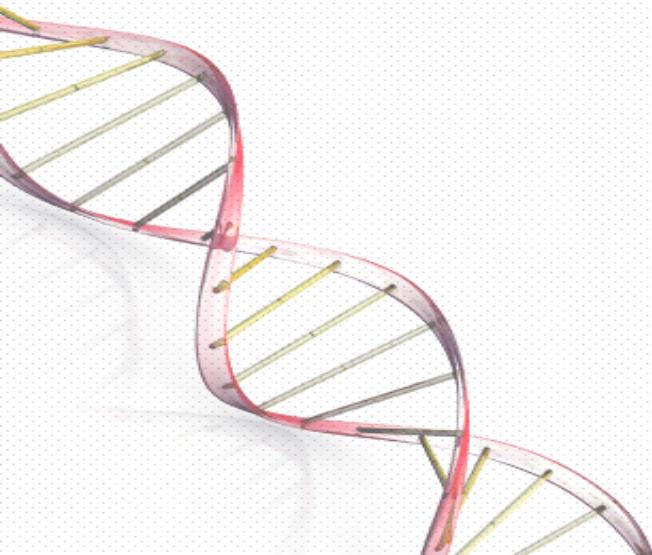
Sequence Analysis
[BLAST / FASTA](#) - Sequence similarity search
[MOTIF](#) - Sequence motif search
[MAFFT / CLUSTALW / PRRN](#) - Multiple alignment
[TREE](#) - Phylogenetic analysis

Genome Analysis

[ViPTree](#) - The Viral Proteomic Tree Server *New!*
[OC Viewer](#) - KEGG ortholog clusters *Updated!*
[REST service](#) is available
[KAAS](#) - KEGG automatic annotation server



DB growth curve

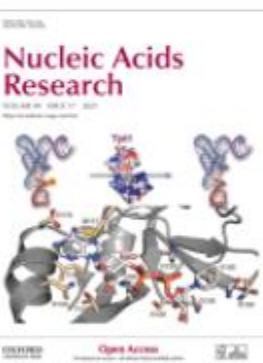


NAR database issue

Nucleic Acids Research

Issues Section browse ▾ Advance articles Submit ▾ Purchase About ▾ All Nucleic Acids Research ▾ Advanced Search

COVID-19 – Support for our author and subscriber community



Latest Issue

Volume 49, Issue 17

7 September 2021

Impact Factor

16.971

5 year Impact Factor

15.542

Senior Executive Editors

Julian Sale

Barry Stoddard



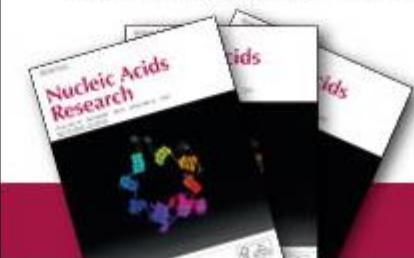
Breakthrough Articles from NAR

Breakthrough Articles present high-impact studies answering long-standing questions in the field of nucleic acids research and/or opening up new areas and mechanistic hypotheses for investigation. These articles are chosen by the Editors on the recommendation of Editorial Board members and Referees.

[Explore the latest breakthrough articles](#)

Explore the 2021 NAR Database Issue

89 new database descriptions

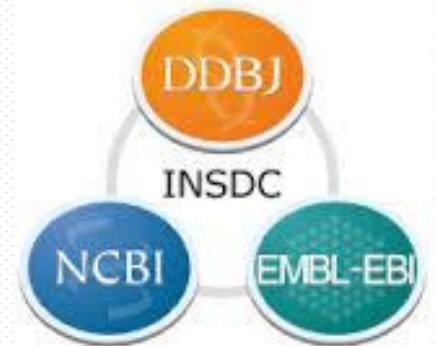
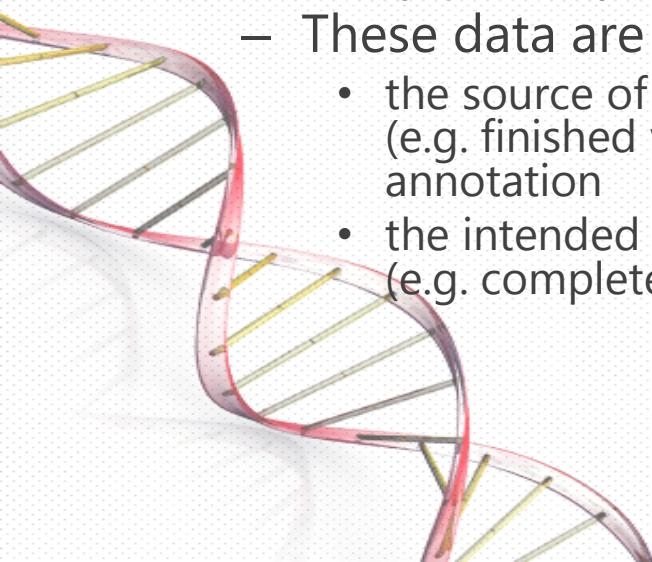


CLICK HERE
TO READ

Advertisemen

Nucleotide databases

- International nucleotide sequence database collaborations
 - Genbank
 - EMBL
 - DDBJ
- The nucleotide sequence databases are data repositories, accepting nucleic acid sequence data from the scientific community and making it freely available.
 - The databases strive for completeness, with the aim of recording every publicly known nucleic acid sequence.
 - These data are heterogeneous, they vary with respect to
 - the source of the material (e.g. genomic versus cDNA), the intended quality (e.g. finished versus single pass sequences), the extent of sequence annotation
 - the intended completeness of the sequence relative to its biological target (e.g. complete versus partial coverage of a gene or a genome).



GenBank entry

Carchesium polypinum HSP70 (HSP70) gene, partial cds

GenBank: AY561304.1

[FASTA](#) [Graphics](#)

Go to:

LOCUS AY561304 441 bp DNA linear INV 26-JUL-2016

DEFINITION Carchesium polypinum HSP70 (HSP70) gene, partial cds.

ACCESSION AY561304

VERSION AY561304.1

KEYWORDS .

SOURCE Carchesium polypinum

ORGANISM [Carchesium polypinum](#)

Eukaryota; Alveolata; Ciliophora; Intramacronucleata;
Oligohymenophorea; Peritrichia; Sessilida; Vorticellidae;
Carchesium.

REFERENCE 1 (bases 1 to 441)

AUTHORS Fu,C., Miao,W. and Shen,Y.

TITLE Phylogeny of Peritrichs Inferred from HSP70 gene

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 441)

AUTHORS Fu,C., Miao,W. and Shen,Y.

TITLE Direct Submission

JOURNAL Submitted (27-FEB-2004) Center of Freshwater Ecology, Institute of Hydrobiology, Chinese Academy of Sciences, Donghu Southern 7, Wuhan, Hubei 430072, P.R.China

FEATURES Location/Qualifiers

source 1..441

/organism="Carchesium polypinum"

/mol_type="genomic DNA"

/db_xref="taxon:168244"

gene <1..>441

/gene="HSP70"

mRNA <1..>441

/gene="HSP70"

CDS /product="HSP70"

<1..>441

ORIGIN

1 ggatttggaaat tgacttggga actaccatac ctttgttagg aattttggcaa aatgaaaagag
61 tttaaaatttat tgctaatgtat caaggaaaca ggactacccc atcttatgttt gcattcaact
121 atactgaaag acttatttggaa gatgctgcta aaaaatcaaac agccaagaat ccaatcaata
181 cagtcttga cgctaagaga cttatttggaa gaaaatttcaa tgatcctatt gtccaaaaag
241 atattaaact ttggccatttc aagggttgaat ctggccctga tgataagccca gtcatttttg
301 tcaaatttaa aggagaaaacc aaaaatgtcc acccagaaga aatctcttca atggttttaa
361 caaaaatgaa ataaaactgcc taatctttct taggataaaac cattaaaaac gctgttgca
421 cagtcccgac ttacttcaac g

//

Customize view

Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

[Find in this Sequence](#)

Related information

[Protein](#)

[Taxonomy](#)

Recent activity

[Turn Off](#) [Clear](#)

Carchesium polypinum HSP70 (HSP70)
gene, partial cds Nucleotide

[Drosophila auraria mRNA for hsp70 heat](#)


[Genomes](#)
[Genome Browser](#)
[Tools](#)
[Mirrors](#)
[Downloads](#)
[My Data](#)
[Help](#)
[About Us](#)


Our story

On June 22, 2000, UCSC and the other members of the International Human Genome Project consortium completed the first working draft of the human genome assembly, forever ensuring free public access to the genome and the information it



What's new

Jul. 14, 2017 - New alternative mouse strain hub and strain-specific annotations

Jul. 14, 2017 - New video: View discontinuous regions together

Search: for

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

Browse a Genome

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Favourite genomes



Human
GRCh38.p10



Mouse
GRCm38.p5



Zebrafish
GRCz10

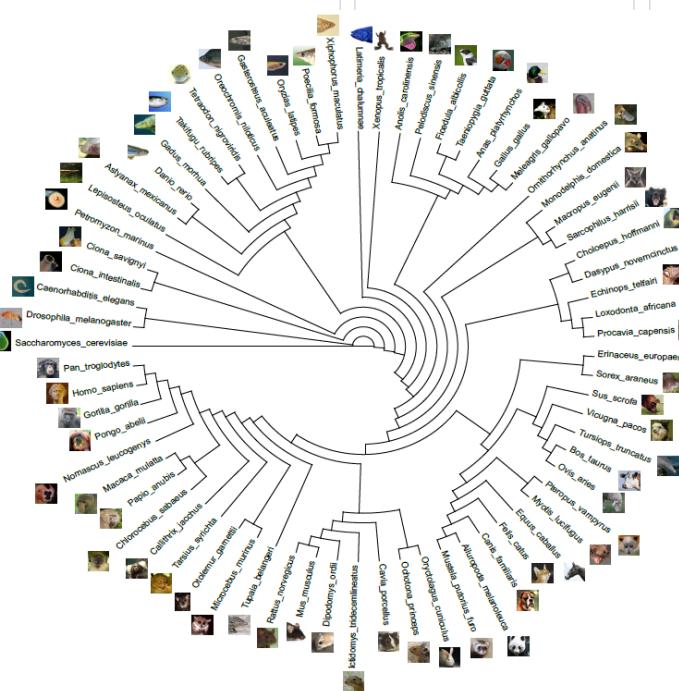
[Edit favourites](#)

All genomes

[-- Select a species --](#)

[View full list of all Ensembl species](#)

Other species are available



Find a Data Display



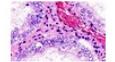
Not sure how to find the data visualisation you need? With our new [Find a Data Display](#) page, you can choose a gene, region or variant and then browse a selection of relevant visualisations

[Try it now!](#)

Variant Effect Predictor



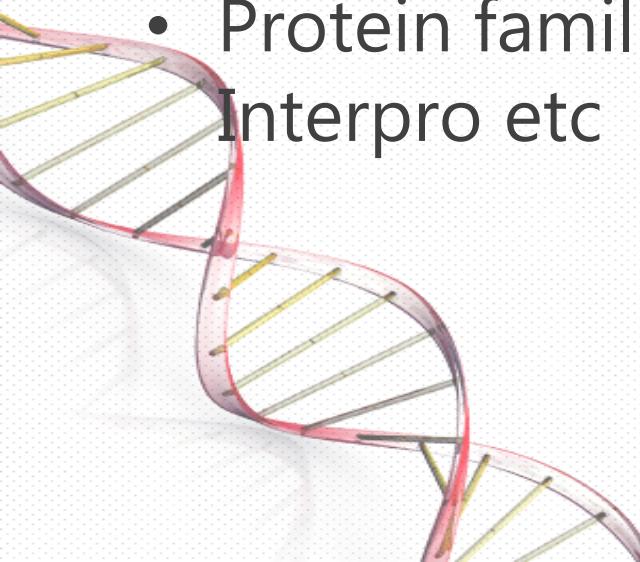
Gene expression in different tissues



[Rare genes across species](#)



[My own data in Ensembl](#)



Protein Databases

- Sequences are in Uniprot
- Structures are in PDB
- Enzyme classifications EC
- Protein families: Pfam, Interpro etc



UniProt: The Universal Protein Resource

The gold-standard, comprehensive resource for protein sequence and functional annotation data.



PRIDE: The Proteomics Identifications Database

An archive of protein expression data determined by mass spectrometry.



InterPro

A database for the classification of proteins into families, domains and conserved sites.



Pfam

A database of hidden Markov models and alignments to describe conserved protein families and domains.



Clustal Omega

Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.



InterProScan 5

InterProScan 5 searches sequences against InterPro's predictive protein signatures. Please note that [InterProScan 4.8 has been retired](#).



Clustal Phylogeny

Commonly used phylogenetic tree generation methods provided by the ClustalW2 program.



NCBI BLAST [protein]

Fast local similarity search tool for protein sequence databases.

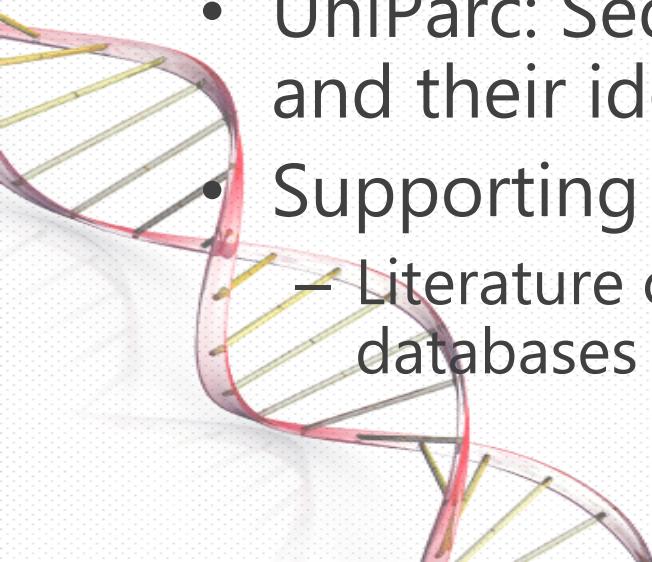


Protein Data Bank in Europe

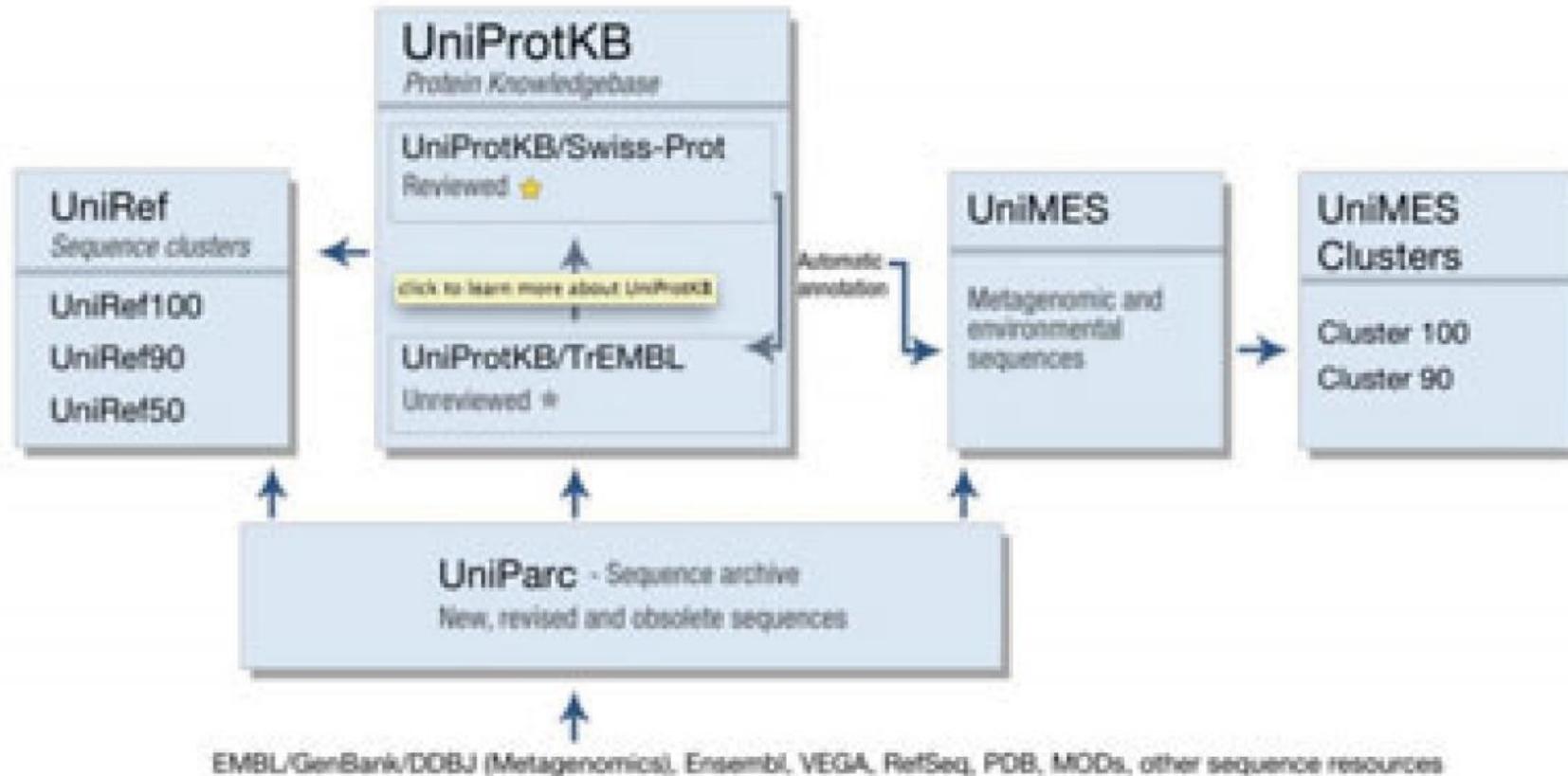
The European resource for the collection, organization and dissemination of 3D structural data (from PDB and EMDB) on biological macromolecules and their complexes.

Uniprot

- UniProtKB: Protein knowledgebase, consists of two sections:
 - Swiss-Prot, which is manually annotated and reviewed.
 - TrEMBL, which is automatically annotated and is **not** reviewed.
- Includes complete and reference proteome sets.
- UniRef: Sequence clusters, used to speed up sequence similarity searches.
- UniParc: Sequence archive, used to keep track of sequences and their identifiers.
- Supporting data
 - Literature citations, keywords, subcellular locations, cross-referenced databases and more.

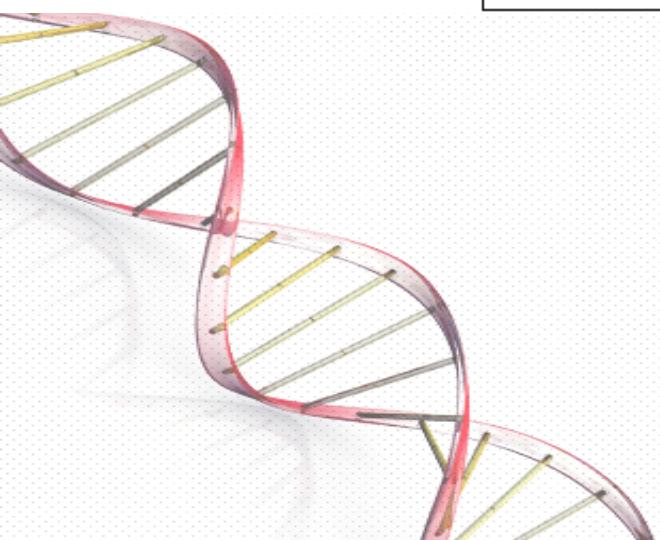
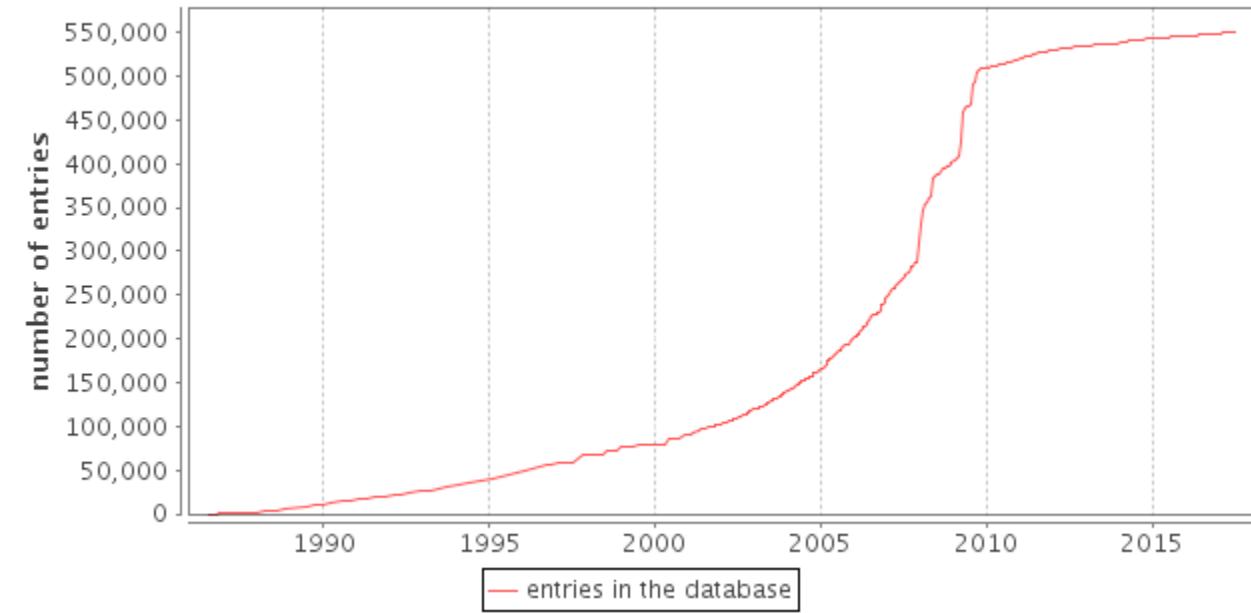


Uniprot

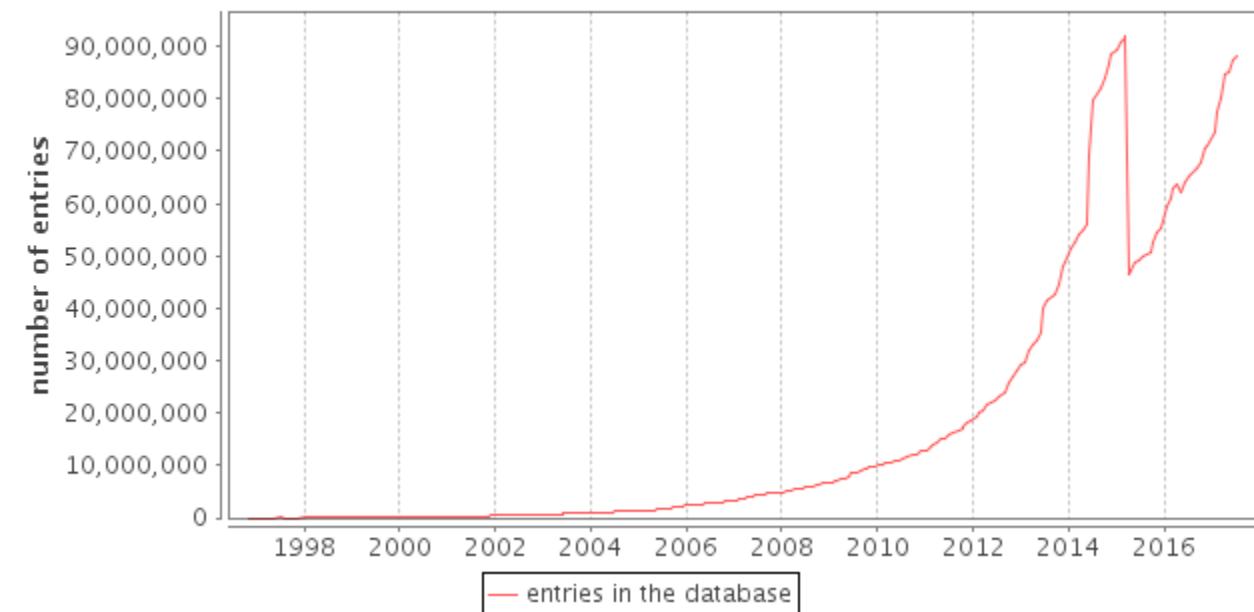


Growth of Protein sequences

Number of entries in UniProtKB/Swiss-Prot over time



Number of entries in UniProtKB/TrEMBL over time



PDB

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

Detailed Select a Report Match score: Higher to Lower Download Files

Homo sapiens (155)
Influenza A virus (48)
Sus scrofa (44)
Rattus norvegicus (42)
Hirudo medicinalis (37)
Escherichia coli (20)
Other (218)

UNIPROT MOLECULE NAME
Cationic trypsin (440)
Pancreatic trypsin inhibitor (110)
Trypsin (60)
Hemagglutinin (50)
Prothrombin (41)
Anionic trypsin-2 (36)
Urokinase-type plasminoge ... (24)
Refine Query

TAXONOMY
Eukaryota (905)
Viruses (76)
Bacteria (73)
Other (7)
Unassigned (1)
Archaea (1)

EXPERIMENTAL METHOD
X-ray (950)

1H4W
Structure of human trypsin IV (brain trypsin)
[Katona, G.](#), [Berglund, G.I.](#), [Hajdu, J.](#), [Graf, L.](#), [Szilagyi, L.](#)
(2002) *J Mol Biol* **315** 1209

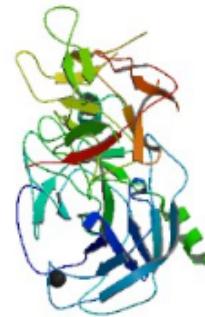
Released: 2/11/2002 **Macromolecule:** TRYPSIN IVA (protein)
Method: X-ray Diffraction **Unique Ligands:** BEN, CA
Resolution: 1.7 Å **Search term match score:** 283.33
Residue Count: 224

Matched fields in 1H4W.cif:

- _citation.title: Crystal Structure Reveals Basis for the Inhibitor Resistance of Human Brain Trypsin.
- _entity.pdbx_description: TRYPSIN IVA, BENZAMIDINE, CALCIUM ION
- _struct.title: Structure of human trypsin IV (brain trypsin)

4AN7
Kunitz type trypsin inhibitor complex with porcine trypsin
[Patil, D.N.](#), [Chaudhary, A.](#), [Sharma, A.K.](#), [Tomar, S.](#), [Kumar, P.](#)
(2012) *FEBS J* **279** 4547

Released: 11/7/2012 **Macromolecule:** TRYPSIN (protein)
Method: X-ray Diffraction **TRYPSIN INHIBITOR (protein)**
Resolution: 2.23 Å



PDB

RCSB PDB Deposit ▾ Search ▾ Visualize ▾ Analyze ▾ Download ▾ Learn ▾ More ▾ MyPDB Login ▾

Advanced Search | Browse by Annotations | Search History (1) | Previous Results (1031)

PDB-101 Worldwide Protein Data Bank EMDDataBank Worldwide Nucleic Acid Database Protein Data Bank Foundation

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment Literature

Biological Assembly 1 1H4W Display Files ▾ Download Files ▾

Structure of human trypsin IV (brain trypsin)
DOI: 10.2210/pdb1h4w/pdb
Classification: HYDROLASE
Deposited: 2001-05-15 Released: 2002-02-11
Deposition author(s): Katona, G., Berglund, G.I., Hajdu, J., Graf, L., Szilagyi, L.
Organism: Homo sapiens
Expression System: ESCHERICHIA COLI
Mutation(s): 1

Experimental Data Snapshot
Method: X-RAY DIFFRACTION
Resolution: 1.7 Å
R-Value Free: 0.203
R-Value Work: 0.188

wwPDB Validation
3D Report Full Report

Metric	Percentile Ranks	Value
Rfree	1	0.217
Clashscore	4	0
Ramachandran outliers	0	0
Sidechain outliers	1.1%	0
RSRZ outliers	0	0

Worse Better
Percentile relative to all X-ray structures
Percentile relative to X-ray structures of similar resolution

Literature Download Primary Citation ▾

Crystal Structure Reveals Basis for the Inhibitor Resistance of Human Brain Trypsin.

View in 3D: NGL or JSmol (in Browser)
Standalone Viewers
Simple Viewer Protein Workshop
Ligand Explorer Kiosk Viewer
Protein Symmetry: Asymmetric (View in 3D)
Protein Stoichiometry: Monomer



Family: P22_Cro (PF14549)

3 architectures

57 sequences

1 interaction

49 species

5 structures

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

Jump to... ↗

enter ID/acc

Go

Summary: DNA-binding transcriptional regulator Cro

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

No Wikipedia article

Pfam

InterPro

This tab holds the annotation information that is stored in the Pfam database. As we move to using Wikipedia as our main source of annotation, the contents of this tab will be gradually replaced by the Wikipedia tab.

DNA-binding transcriptional regulator Cro [Provide feedback](#)

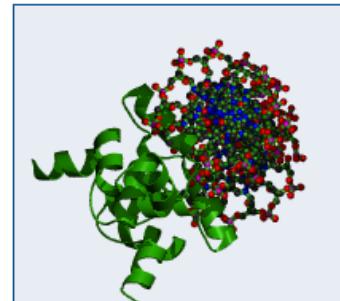
Bacteriophage P22 Cro protein represses genes normally expressed in early phage development and is necessary for the late stage of lytic growth. It does this by binding to the OL and OR operator-regions normally used by the repressor protein for lysogenic maintenance.

Literature references

1. Poteete AR, Hehir K, Sauer RT;; Biochemistry. 1986;25:251-256.: Bacteriophage P22 Cro protein: sequence, purification, and properties. [PUBMED:3954988](#) [EPMC:3954988](#)
2. Newlove T, Konieczka JH, Cordes MH;; Structure. 2004;12:569-581.: Secondary structure switching in Cro protein evolution. [PUBMED:15062080](#) [EPMC:15062080](#)
3. Dubrava MS, Ingram WM, Roberts SA, Weichsel A, Montfort WR, Cordes MH;; Protein Sci. 2008;17:803-812.: N15 Cro and lambda Cro: orthologous DNA-binding domains with completely different but equally effective homodimer interfaces. [PUBMED:18369196](#) [EPMC:18369196](#)

Internal database links

SCOOP:

[HTH_1](#) [HTH_11](#) [HTH_19](#) [HTH_23](#) [HTH_24](#) [HTH_26](#) [HTH_28](#)
[HTH_3](#) [HTH_31](#) [Phage_C1_repr](#) [Sigma70_r4](#)
[YdaS_antitoxin](#)Similarity to PfamA
using HHSearch:[HTH_1](#) [HTH_3](#) [YdaS_antitoxin](#)[Example structure](#)

[PDB entry 3QWS](#): Structure of N15 Cro complexed with consensus operator DNA
[View a different structure](#):

3QWS ▾

CATH (and related databases such as SCOP, ECOD)



CATH / Gene3D v4.1

53 million protein domains classified into

Search by keywords, PDB code

 **3D Structure**
Find out what 3D structure your protein adopts

[Find out more](#) [Go](#)

 **Conserved Sites**
Look at protein sites that are highly conserved and implicated in function

[Find out more](#) [Go](#)

Top of CATH Hierarchy (4 Classes)

- **C 1** Mainly Alpha
 - ▷ **A 1.10** Orthogonal Bundle
 - ▷ **A 1.20** Up-down Bundle
 - ▷ **A 1.25** Alpha Horseshoe
 - ▷ **A 1.40** Alpha solenoid
 - ▷ **A 1.50** Alpha/alpha barrel
 - ▷ **C 2** Mainly Beta
 - **C 3** Alpha Beta
 - ▷ **A 3.10** Roll
 - ▷ **A 3.15** Super Roll
 - ▷ **A 3.20** Alpha-Beta Barrel
 - ▷ **A 3.30** 2-Layer Sandwich
 - ▷ **A 3.40** 3-Layer(aba) Sandwich
 - ▷ **A 3.50** 3-Layer(bba) Sandwich
 - ▷ **A 3.55** 3-Layer(bab) Sandwich
 - ▷ **A 3.60** 4-Layer Sandwich
 - ▷ **A 3.65** Alpha-beta prism
 - ▷ **A 3.70** Box
 - ▷ **A 3.75** 5-stranded Propeller
 - ▷ **A 3.80** Alpha-Beta Horseshoe
 - ▷ **A 3.90** Alpha-Beta Complex
 - ▷ **A 3.100** Ribosomal Protein L15; Chain: K; domain 2
 - ▷ **C 4** Few Secondary Structures
- 5 Architectures, 396 Folds, 908 Superfamilies, 61579 Domains
287 Folds, 545 Superfamilies, 43441 Domains
101 Folds, 330 Superfamilies, 15652 Domains
5 Folds, 27 Superfamilies, 1714 Domains
1 Folds, 1 Superfamilies, 11 Domains
2 Folds, 5 Superfamilies, 761 Domains
20 Architectures, 241 Folds, 547 Superfamilies, 78049 Domains
14 Architectures, 628 Folds, 1160 Superfamilies, 165745 Domains
58 Folds, 101 Superfamilies, 12126 Domains
3 Folds, 3 Superfamilies, 12 Domains
18 Folds, 47 Superfamilies, 13542 Domains
224 Folds, 495 Superfamilies, 46678 Domains
126 Folds, 287 Superfamilies, 65246 Domains
11 Folds, 17 Superfamilies, 3194 Domains
6 Folds, 6 Superfamilies, 28 Domains
16 Folds, 18 Superfamilies, 4785 Domains
1 Folds, 2 Superfamilies, 436 Domains
1 Folds, 1 Superfamilies, 186 Domains
1 Folds, 2 Superfamilies, 143 Domains
3 Folds, 4 Superfamilies, 386 Domains
159 Folds, 176 Superfamilies, 18797 Domains
1 Folds, 1 Superfamilies, 186 Domains
1 Architectures, 108 Folds, 122 Superfamilies, 3626 Domains

Functional Classification

- EC Classifications

CLASSIFICATION OF ENZYMES		
Group of Enzyme	Reaction Catalysed	Examples
1. Oxdoreductases	Transfer of hydrogen and oxygen atoms or electrons from one substrate to another.	Dehydrogenases Oxidases
2. Transferases	Transfer of a specific group (a phosphate or methyl etc.) from one substrate to another.	Transaminase Kinases
3. Hydrolases	Hydrolysis of a substrate.	Estrases Digestive enzymes
4. Isomerases	Change of the molecular form of the substrate.	Phospho hexo isomerase, Fumarase
5. Lyases	Nonhydrolytic removal of a group or addition of a group to a substrate.	Decarboxylases Aldolases
6. Ligases (Synthetases)	Joining of two molecules by the formation of new bonds.	Citric acid synthetase

A

CSA LITERATURE entry for 1q6x	
E.C. name	choline O-acetyltransferase
Species	<i>Rattus norvegicus</i> (Rat)
E.C. Number	2.3.1.6
CSA Homologues of 1q6x	There are 24 Homologs
CSA Entries With UniProtID	P32738
PDBe Entry	1q6x
PDBSum Entry	1q6x
MACIE Entry	1q6x

B

Show Site 1

Options Antialias Rotation

Color: Default

Style: Cartoon

Surface: None

C

Literature Report

Introduction	Choline acetyltransferase synthesises the neurotransmitter acetylcholine from choline in neurones and other cell types. It catalyses the reversible transfer of an acetyl group between acetyl CoA and choline, and belongs to the choline/carnitine acyltransferase family which also includes enzymes involved in fatty acid metabolism.
Mechanism	His 334 acts as a general base to remove the proton from the choline OH group as the oxygen attacks the carbonyl of acetyl CoA. The resulting tetrahedral oxyanion intermediate is stabilised by Ser 550. Collapse of the tetrahedral intermediate releases CoA which is protonated by the His 334. Tyr 95 and Pro 108 function to stabilise the unprotonated form of His 334 N-epsilon so that it can act as a general base in the first step of the reaction.
Reaction	

D

Catalytic Sites for 1q6x

Annotated By Reference To The Literature - Site 1

Residue	Chain	Number	UniProtKB Number	Functional Part	Function	Target	Description
His	A	334	334	Side Chain	Acid / Base	Substrate	Acts as a general base, extracting a proton from the attacking hydroxyl group of choline. Later protonates the departing sulphhydryl group of CoA.

Function classifications: GO



Gene Ontology Consortium

[Home](#) [Documentation](#) [Downloads](#) [Community](#) [Tools](#) [About](#)

Enrichment analysis

Your gene IDs here...

biological process ▾

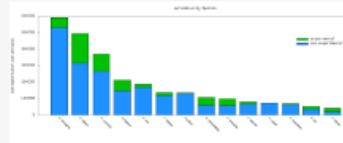
Homo sapiens ▾

Submit

Help

Powered by [PANTHER](#)

Statistics



Other GOC tools

Explore other GOC tools in the AmiGO software suite

Gene Ontology Consortium

Search GO data

Search for terms and gene products...

Search

Ontology

[Filter classes](#)

[Download ontology](#)

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

molecular function

molecular activities of gene products

cellular component

where gene products are active

biological process

pathways and larger processes made up of the activities of multiple gene products.

[more](#)

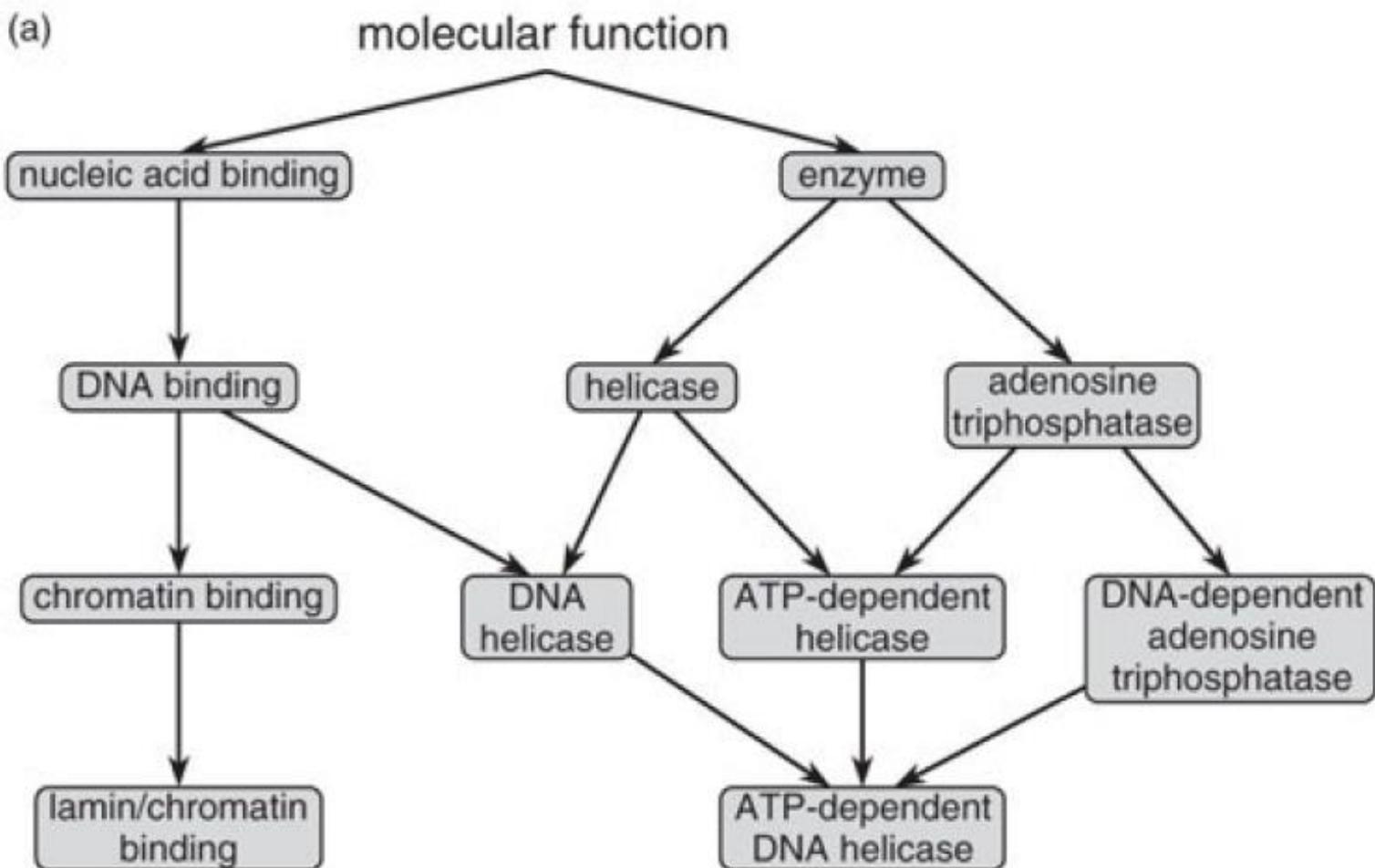
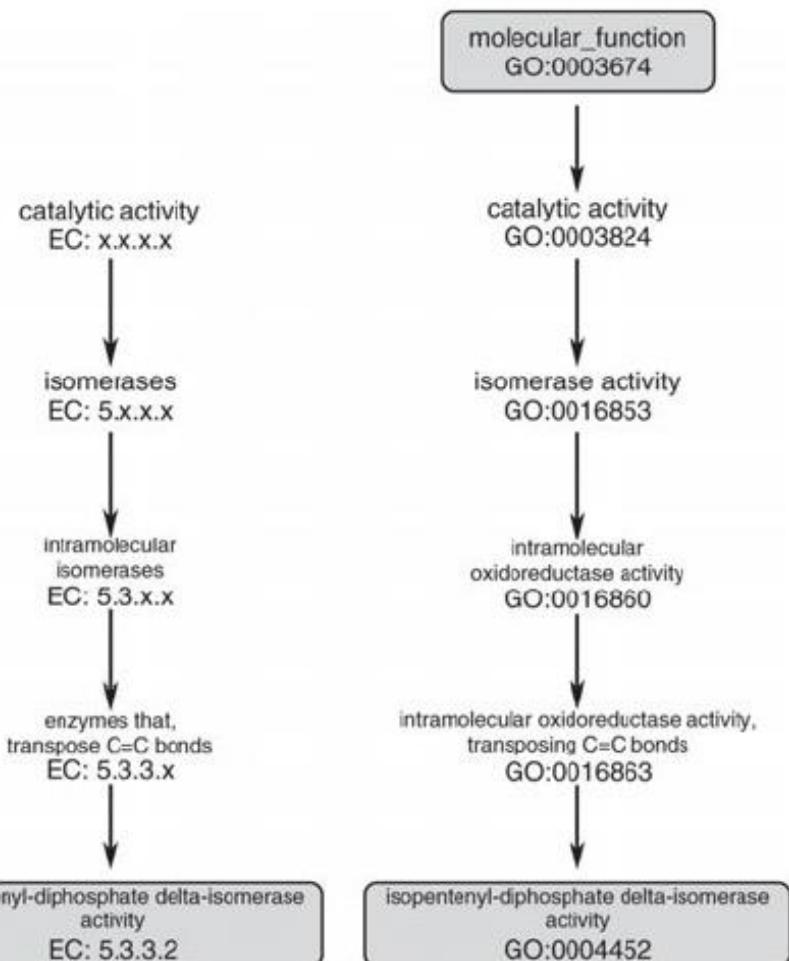
Annotations

[Download annotations \(standard files\)](#)

[Filter and download \(customizable files <100k lines\)](#)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. [more](#)

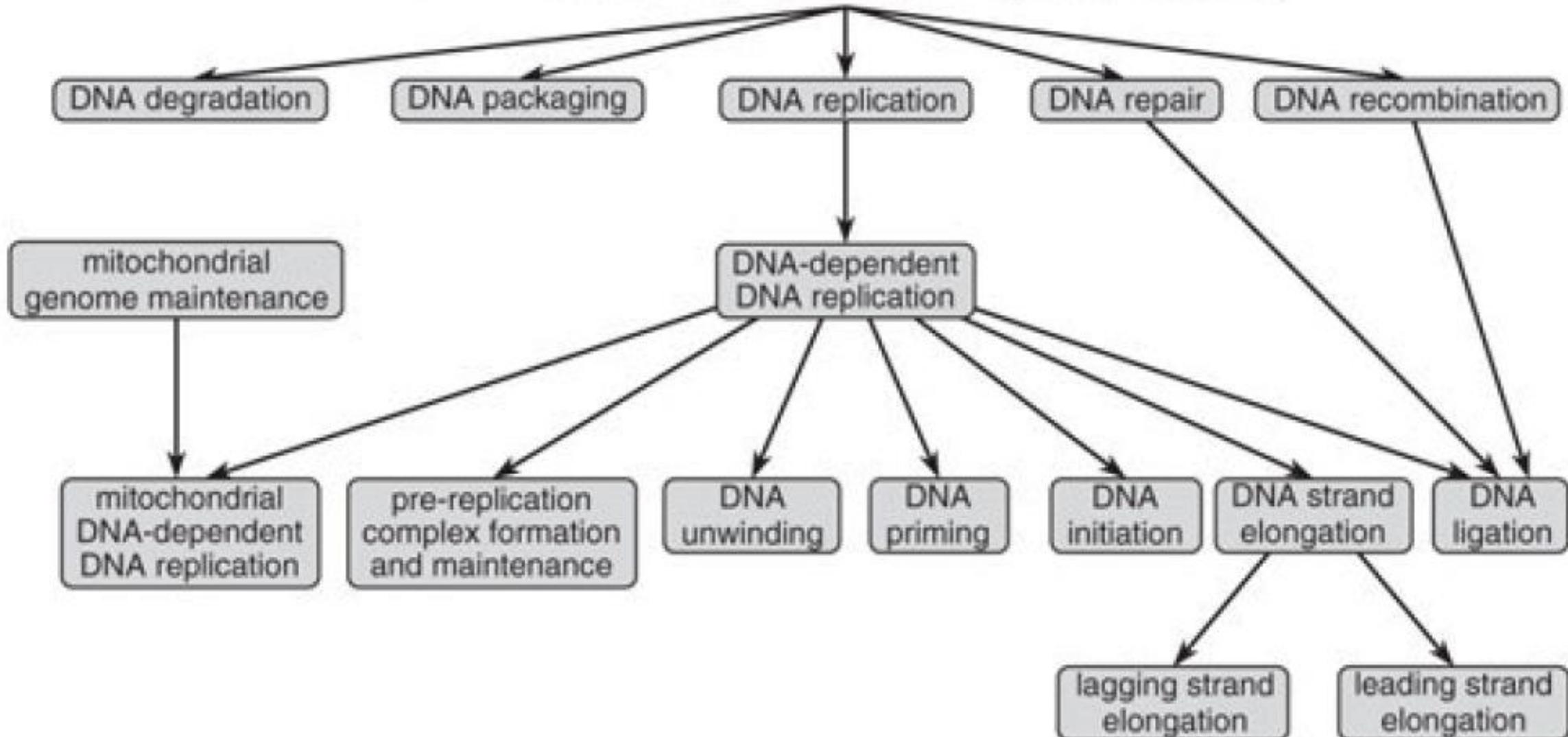
GO - molecular function



GO metabolic process

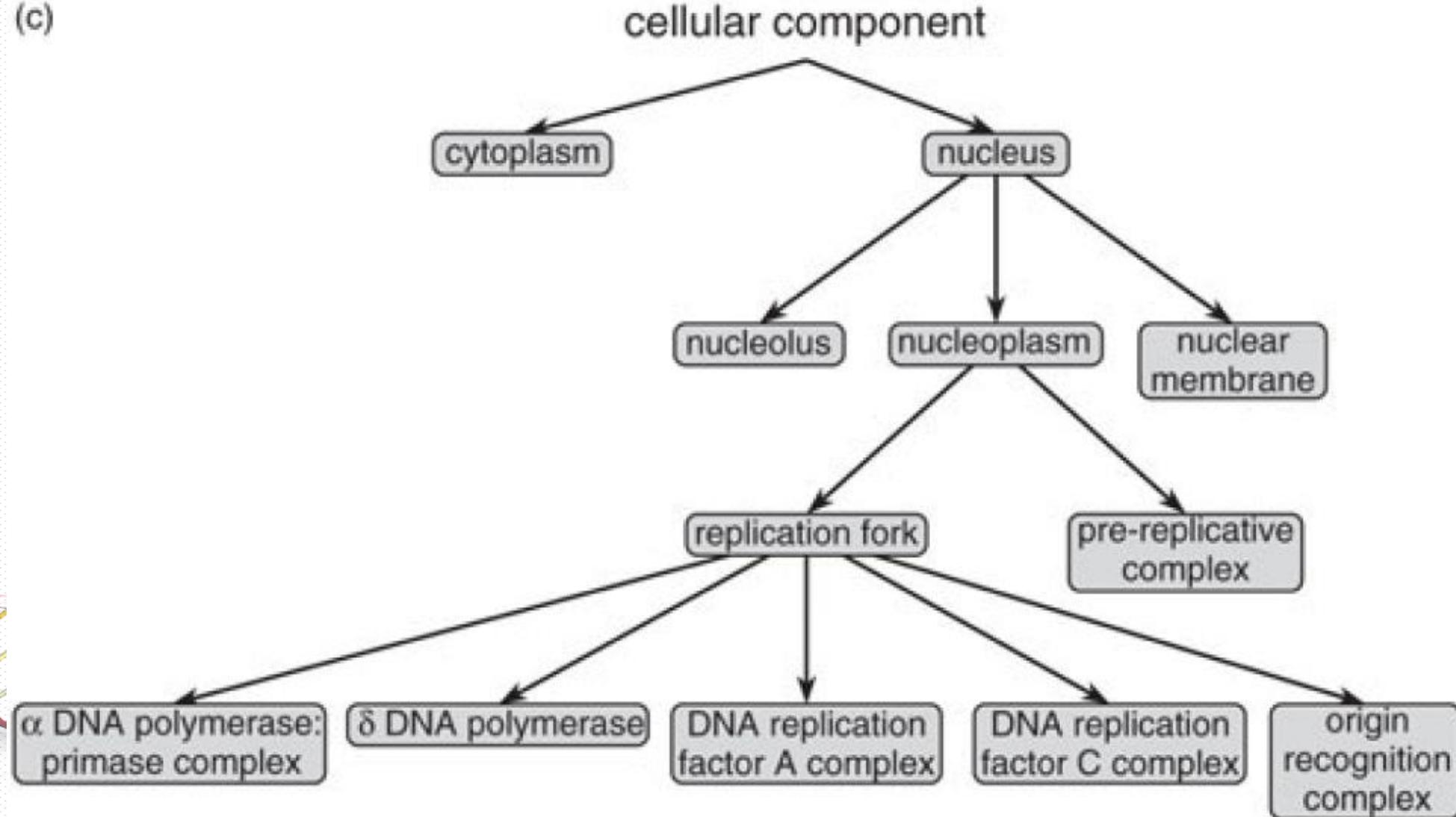
(b)

DNA metabolic process (biological process)



GO cellular component

(c)



Expression databases

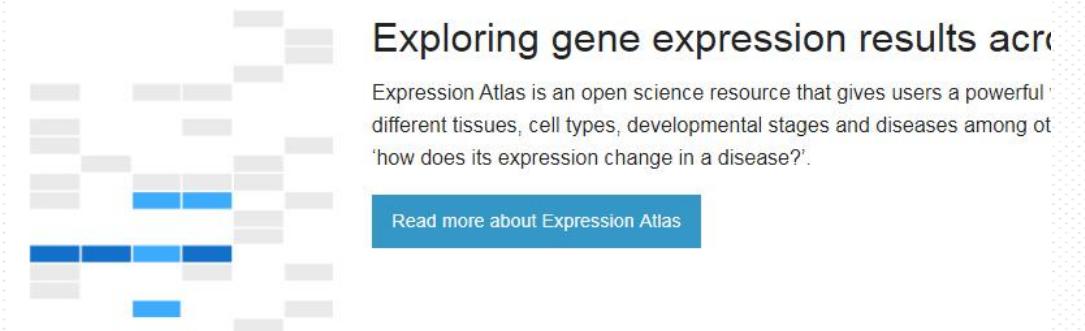
- Gene Expression, GeneExpressAtlas
 - total experiments 3476
 - total genes 692307
- Protein Expression: Pride
 - 25,816 Experiments
 - 11,042,722 Identified Proteins

EMBL-EBI Services Research Training About us

Expression Atlas

Gene expression across species and biological conditions

Home Download Release notes FAQ Help Licence About



Exploring gene expression results across species and conditions

Expression Atlas is an open science resource that gives users a powerful way to explore gene expression data from across many different species, tissues, cell types, developmental stages and diseases among other things. You can ask questions such as 'which genes are expressed in a particular tissue?' or 'how does its expression change in a disease?'.

[Read more about Expression Atlas](#)

PRIDE Archive

Home Submit data Browse data Help Publications About PRIDE Archive

PRIDE > Archive > Search results

Search results

1616 Search results + 1 filters

Filter your results

Field Species Contains -Any- Add filter

Current active filters

Remove all

Species: Homo sapiens (Human)

PXD005823 Comparative proteomic analysis and bioluminescent reporter gene assays to investigate effects of simulated microgravity on Caco-2 cells
Species: [Homo sapiens \(Human\)](#)
Project description: Microgravity is one of the most important features in spaceflight ([More](#))
Made public: 2017-07-26
 [Biological Dataset](#)

PXD006739 Quantitative assessment of the effects of trypsin digestion methods on immunoprecipitation-mass spectrometry-based protein-protein interaction analysis - part2
Species: [Homo sapiens \(Human\)](#)
Project description: Immunoprecipitation-mass spectrometry (IP-MS) has become the method of choice for protein interaction analysis ([More](#))
Made public: 2017-07-26
 [Technical Dataset](#)

PXD005984 Quantitative assessment of the effects of trypsin digestion methods on immunoprecipitation-mass spectrometry-based protein-protein interaction analysis
Species: [Homo sapiens \(Human\)](#)

Molecular Interactions

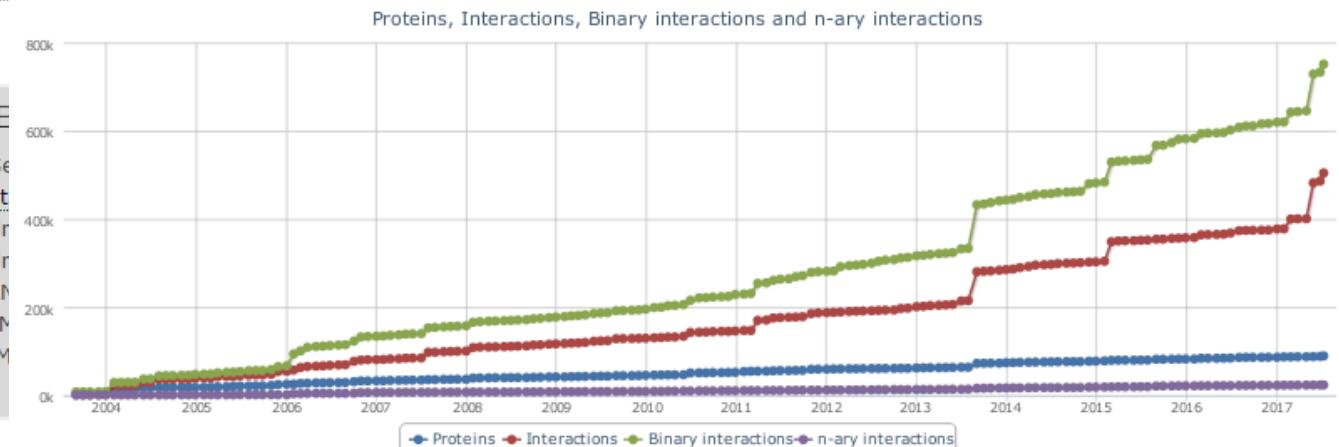
IntAct Molecular Interaction Database

IntAct provides a freely available, open source database system and analysis tools for molecular interaction data. All interactions are derived from literature or direct user submissions and are freely available. The IntAct Team also produce the [Complex Portal](#).

Search in IntAct

[Search](#) [Search Tips](#)

- Gene
- Structure
- UniProt
- UniRef
- RNA
- PDB
- IMEx



Data Content

- Publications: **14689**
- Interactions: **751286**
- Interactors: **99807**

Citing IntAct

The MIntAct project--
IntAct as a common
curation platform for
11 molecular
interaction databases.

Orchard S et al

[PMID: 24234451]

[\[Full Text\]](#)

Submission

Submit your data to
IntAct to increase its
visibility and usability!

Training

Online & upcoming courses

Contributors

Manually curated content is added to IntAct by curators at EBI and the following organisations:



HINT Home Download FAQ

The HINT homepage features a prominent "HINT" logo and the tagline "High-quality Interactomes". It includes links for "Home", "Download", and "FAQ". A large search bar at the bottom allows users to enter protein names separated by semicolons, commas, or spaces. Below the search bar are dropdown menus for "H. sapiens", "Binary", and "High quality". A "Submit" button is located at the bottom right. The background of the page is a light blue network graph with nodes and connections.

Reactions and Pathways



REACTOME

A CURATED PATHWAY DATABASE

About Content Documentation Tools Community Download Contact e.g. O95631, NTN1, signaling Search

Browse Pathways Analyze Data Reactome FlViz app

User Guide Data Download Contact Us

About Reactome

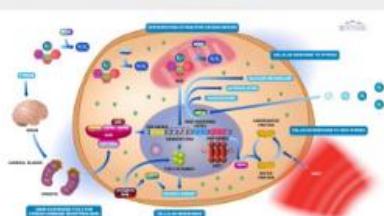
Reactome is a free, open-source, curated and peer reviewed pathway database. Our goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education. The current version (v61) of Reactome was released on June 22, 2017.

If you use Reactome in Asia, we suggest using our Chinese mirror site at reactome.ncpsb.org.

OICR NYU Langone CSH EMBL-EBI

Tweets

Current Version: Reactome V61

reactome @reactome How do our cells respond to stress? Explore it using our #PathwayBrowser starting at the #interactive #illustration reactome.org/PathwayBrowser... 

reactome Retweeted Embed View on Twitter



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions and relations

Menu PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG

Select prefix map Organism Enter keywords Help

[New pathway maps | Update history]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on molecular interaction, reaction and relation networks for:

1. Metabolism

Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycogen Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure

2. Genetic Information Processing

3. Environmental Information Processing

4. Cellular Processes

5. Organismal Systems

6. Human Diseases

7. Drug Development

KEGG PATHWAY is a reference database for **Pathway Mapping**.

Pathway Identifiers

Each pathway map is identified by the combination of 2-4 letter prefix code and 5 digit number (see [KEGG Identifier](#)). The prefix has the following meaning:

map manually drawn pathway

Entrez

Search NCBI databases

Help

Literature

Books	books and reports
MeSH	ontology used for PubMed indexing
NLM Catalog	books, journals and more in the NLM Collections
PubMed	scientific & medical abstracts/citations
PubMed Central	full-text journal articles

Health

ClinVar	human variations of clinical significance
dbGaP	genotype/phenotype interaction studies
GTR	genetic testing registry
MedGen	medical genetics literature and links
OMIM	online mendelian inheritance in man
PubMed Health	clinical effectiveness, disease and drug reports

Genomes

Assembly	genome assembly information
BioCollections	museum, herbaria, and other biorepository collections
BioProject	biological projects providing data to NCBI
BioSample	descriptions of biological source materials
Clone	genomic and cDNA clones
dbVar	genome structural variation studies
Genome	genome sequencing projects by organism
GSS	genome survey sequences
Nucleotide	DNA and RNA sequences
Probe	sequence-based probes and primers
SNP	short genetic variations
SRP	high-throughput DNA and RNA sequence read archive

Genes

EST	expressed sequence tag sequences
Gene	collected information about gene loci
GEO DataSets	functional genomics studies
GEO Profiles	gene expression and molecular abundance profiles
HomoloGene	homologous gene sets for selected organisms
PopSet	sequence sets from phylogenetic and population studies
UniGene	clusters of expressed transcripts

Proteins

Conserved Domains	conserved protein domains
Protein	protein sequences
Protein Clusters	sequence similarity-based protein clusters
Structure	experimentally-determined biomolecular structures

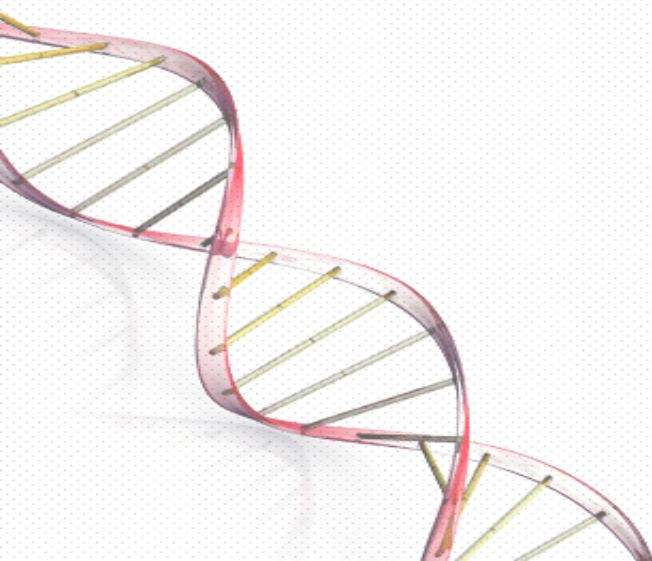
Chemicals

BioSystems	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	bioactivity screening studies
PubChem Compound	chemical information with structures, information and links
PubChem Substance	deposited substance and chemical information



Assignment

- Will be posted online today



References

- Lecture notes of Colin Dewey @ University of Wisconsin-Madison
- Lecture notes of Arne Elofsson @ Stockholm University
- Lecture notes of Yuzhen Ye @ Indiana University

