

The background features a collage of various icons related to biology and bioinformatics. At the top left is a barcode. Below it is a DNA double helix. To the right is a small icon of a document or folder. In the center, there's a complex molecular structure. Below that, a circular diagram represents the cell cycle with phases G1, S, and G2. At the bottom, there are icons of a bacterium, a plant, and a human silhouette.

INTRODUCTION TO BIOINFORMATICS

Please take the initial BIOINF525 questionnaire:
< <http://tinyurl.com/bioinf525-questions> >

Barry Grant
University of Michigan
www.thegrantlab.org



Barry Grant, Ph.D.
bjgrant@umich.edu



Ryan Mills, Ph.D.
remills@umich.edu



Hongyang Li (GSI)
hyangl@umich.edu

COURSE LOGISTICS

Lectures: Tuesdays 2:30-4:00 PM
Rm. 2062 Palmer Commons

Labs: Session I: Thursdays 2:30 - 4:00 PM
Session II: Fridays 10:30 - 12:00 PM
Rm. 2036 Palmer Commons

Website: <http://tinyurl.com/bioinf525-w16>
Lecture, lab and background reading material
plus homework and course announcements

MODULE OVERVIEW

Objective: Provide an introduction to the practice of bioinformatics as well as a practical guide to using common bioinformatics databases and algorithms

1.1. ▶ *Introduction to Bioinformatics*

1.2. ▶ *Sequence Alignment and Database Searching*

1.3 ▶ *Structural Bioinformatics*

1.4 ▶ *Genome Informatics: High Throughput Sequencing Applications and Analytical Methods*

TODAYS MENU

Overview of bioinformatics

- The what, why and how of bioinformatics?
- Major bioinformatics research areas.
- Skepticism and common problems with bioinformatics.




Bioinformatics databases and associated tools

- Primary, secondary and composite databases.
 - Nucleotide sequence databases (GenBank & RefSeq).
 - Protein sequence database (UniProt).
 - Composite databases (PFAM & OMIM).

Database usage vignette

- Searching with ENTREZ and BLAST.
- Reference slides and handout on major databases.

HOMEWORK

-  Complete the **initial course questionnaire**:
<http://tinyurl.com/bioinf525-questions>
-  Check out the “**Background Reading**” material on Ctools:
<http://tinyurl.com/bioinf525-w16>
-  Complete the **lecture 1.1 homework questions**:
<http://tinyurl.com/bioinf525-quiz1>

Q. What is Bioinformatics?

Q. What is Bioinformatics?

“Bioinformatics is the application of computers to the collection, archiving, organization, and analysis of biological data.”

[After Orengo, 2003]

- ... Bioinformatics is a hybrid of biology and computer science
- ... **Bioinformatics is computer aided biology!**

Computer based management and analysis of biological and biomedical data with useful applications in many disciplines, particularly genomics, proteomics, metabolomics, etc...

MORE DEFINITIONS

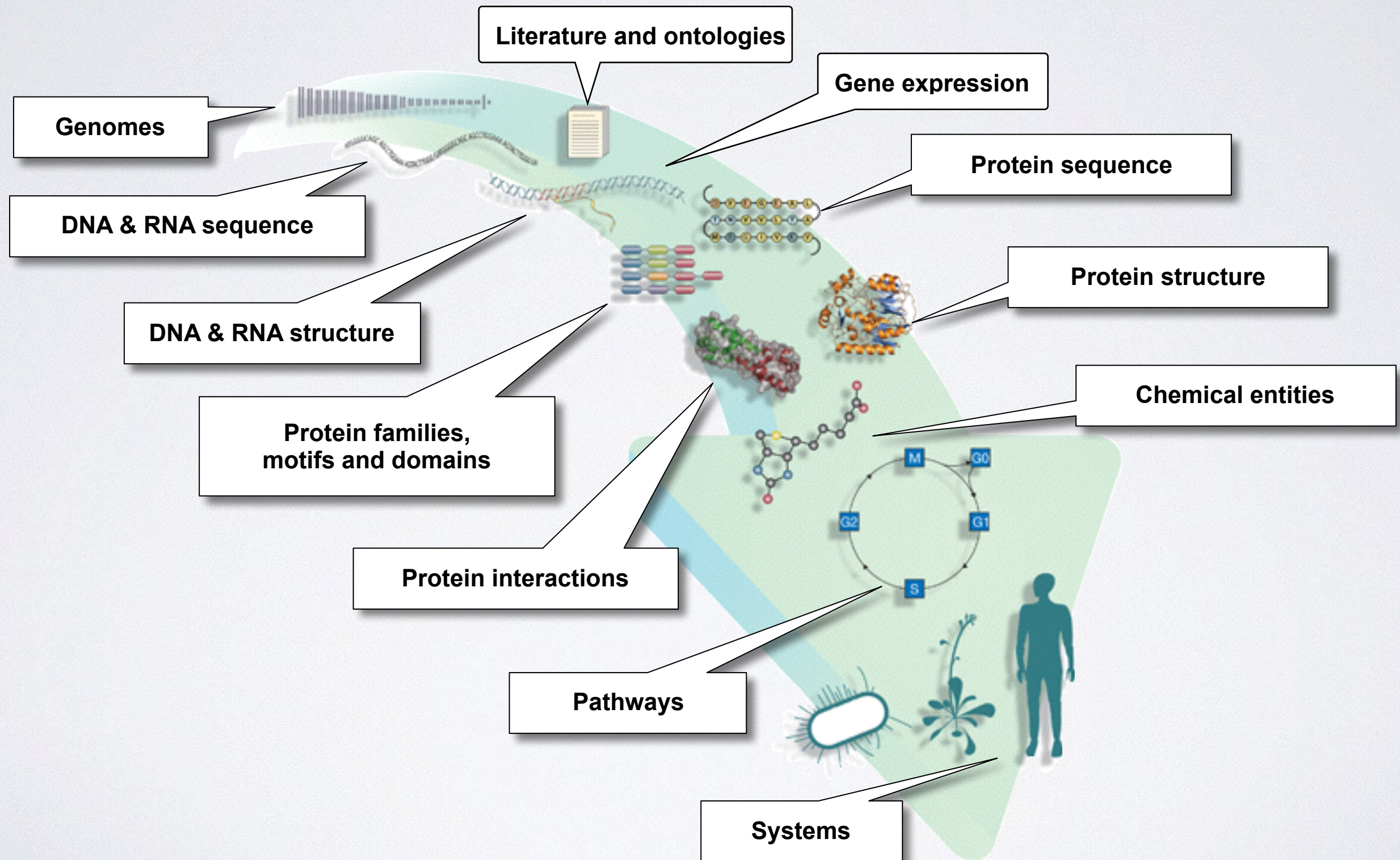
- ▶ “Bioinformatics is conceptualizing biology in terms of **macromolecules** and then applying “**informatics**” **techniques** (derived from disciplines such as applied maths, computer science, and statistics) to **understand** and **organize** the information associated with these molecules, on a **large-scale**.
Luscombe NM, et al. Methods Inf Med. 2001;40:346.
- ▶ “Bioinformatics is research, development, or application of **computational approaches** for expanding the use of **biological, medical, behavioral or health data**, including those to **acquire, store, organize** and **analyze** such data.”
National Institutes of Health (NIH) (<http://tinyurl.com/l3gxr6b>)

MORE DEFINITIONS

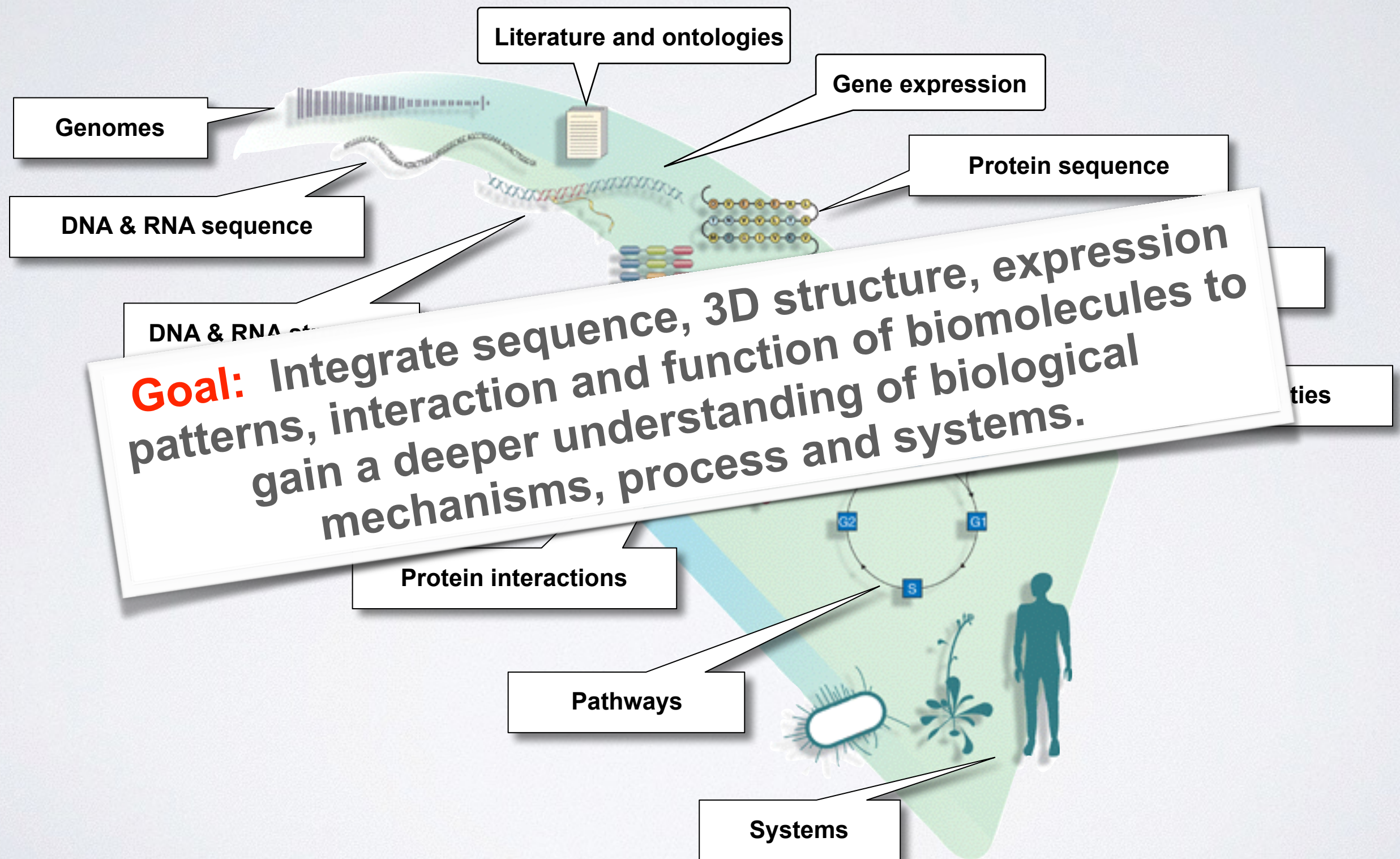
- ▶ “Bioinformatics is conceptualizing biology in terms of **macromolecules** and then applying “informatics” techniques (derived from disciplines such as applied mathematics, computer science, and statistics) to **understand and organize** the information associated with these systems, on a **large-scale**.
Luscombe NM, et al. Methods 2001;40:346.
- ▶ “Bioinformatics is the research, development, or application of **computational approaches** for expanding the use of biological, medical, behavioral or **health data**, including those to **acquire, store, organize and analyze** such data.”
National Institutes of Health (NIH) (<http://tinyurl.com/l3gxr6b>)

Key Point: Bioinformatics is Computer Aided Biology

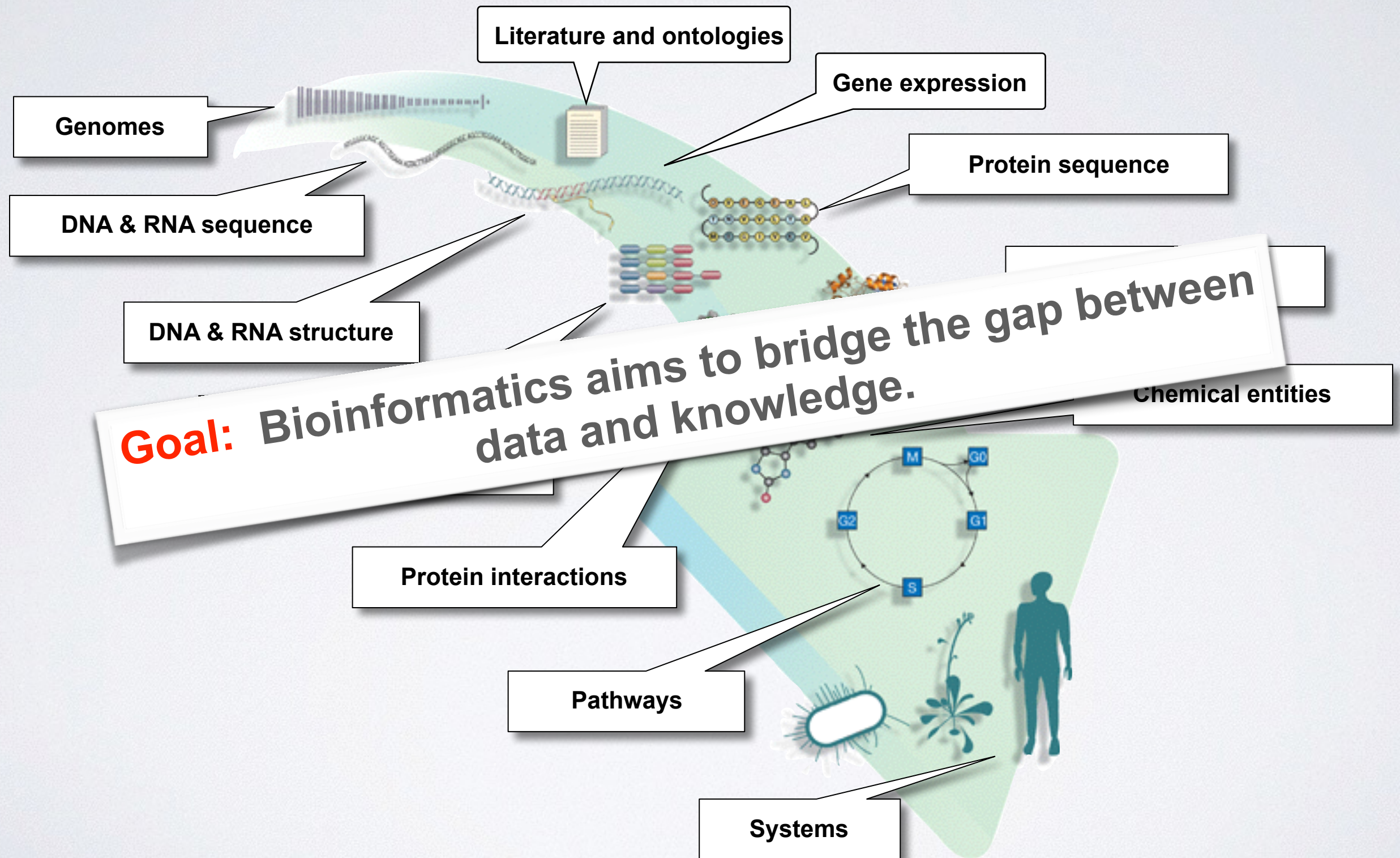
Major types of Bioinformatics Data



Major types of Bioinformatics Data



Major types of Bioinformatics Data



BIOINFORMATICS RESEARCH AREAS

Include but are not limited to:

- Organization, classification, dissemination and analysis of biological and biomedical data (particularly '-omics' data).
- Biological sequence analysis and phylogenetics.
- Genome organization and evolution.
- Regulation of gene expression and epigenetics.
- Biological pathways and networks in healthy & disease states.
- Protein structure prediction from sequence.
- Modeling and prediction of the biophysical properties of biomolecules for binding prediction and drug design.
- Design of biomolecular structure and function.

With applications to Biology, Medicine, Agriculture and Industry

Where did bioinformatics come from?

Bioinformatics arose as molecular biology began to be transformed by the emergence of molecular sequence and structural data

Recap: The key dogmas of molecular biology

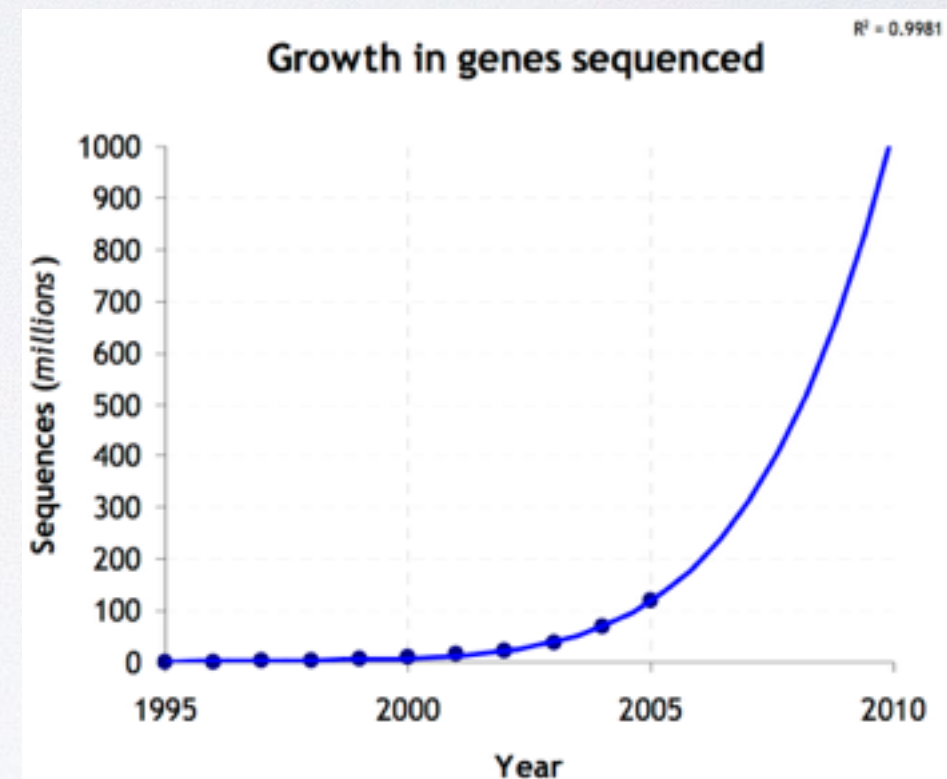
- *DNA sequence determines protein sequence.*
- *Protein sequence determines protein structure.*
- *Protein structure determines protein function.*
- *Regulatory mechanisms (e.g. gene expression) determine the amount of a particular function in space and time.*

Bioinformatics is now essential for the archiving, organization and analysis of data related to these processes.

Why do we need Bioinformatics?

Bioinformatics is necessitated by the rapidly expanding quantities and complexity of biomolecular data

- Bioinformatics provides methods for the efficient:
 - ▶ **storage**
 - ▶ **annotation**
 - ▶ **search and retrieval**
 - ▶ **data integration**
 - ▶ **data mining and analysis**

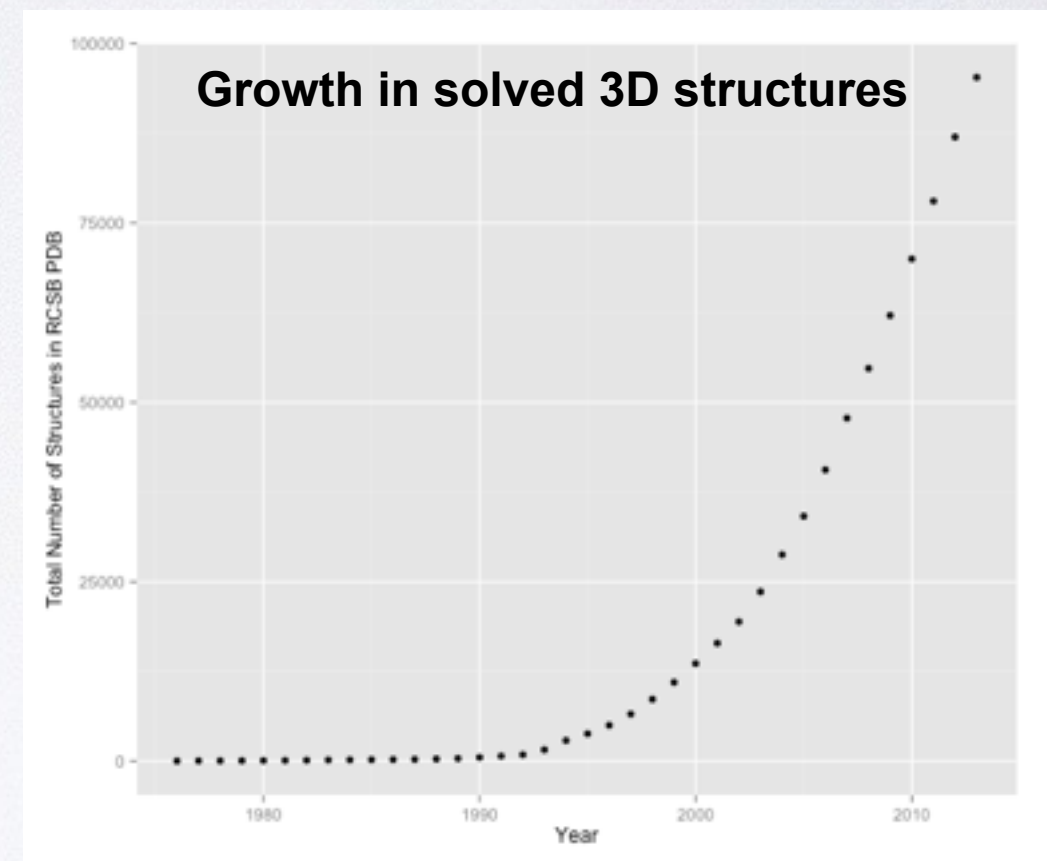


Bioinformatics is essential for the archiving, organization and analysis of data from sequencing, structural genomics, microarrays, proteomics and new high throughput assays.

Why do we need Bioinformatics?

Bioinformatics is necessitated by the rapidly expanding quantities and complexity of biomolecular data

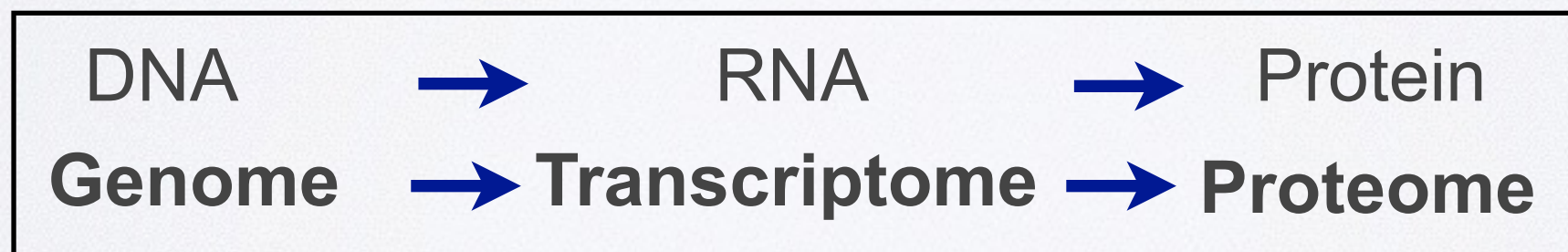
- Bioinformatics provides methods for the efficient:
 - ▶ **storage**
 - ▶ **annotation**
 - ▶ **search and retrieval**
 - ▶ **data integration**
 - ▶ **data mining and analysis**



Bioinformatics is essential for the archiving, organization and analysis of data from sequencing, structural genomics, microarrays, proteomics and new high throughput assays.

How do we do Bioinformatics?

- A “*bioinformatics approach*” involves the application of **computer algorithms**, **computer models** and **computer databases** with the broad goal of understanding the action of both individual genes, transcripts, proteins and large collections of these entities.



How do we actually do Bioinformatics?

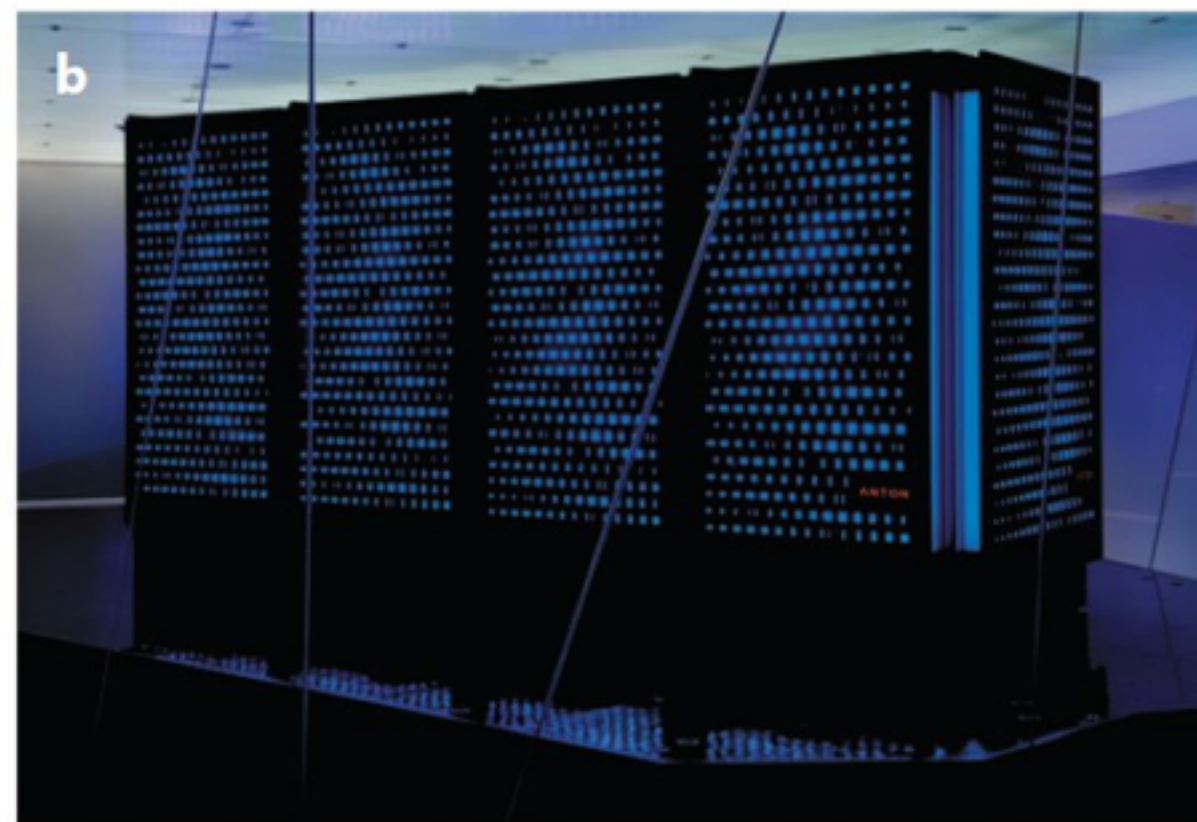
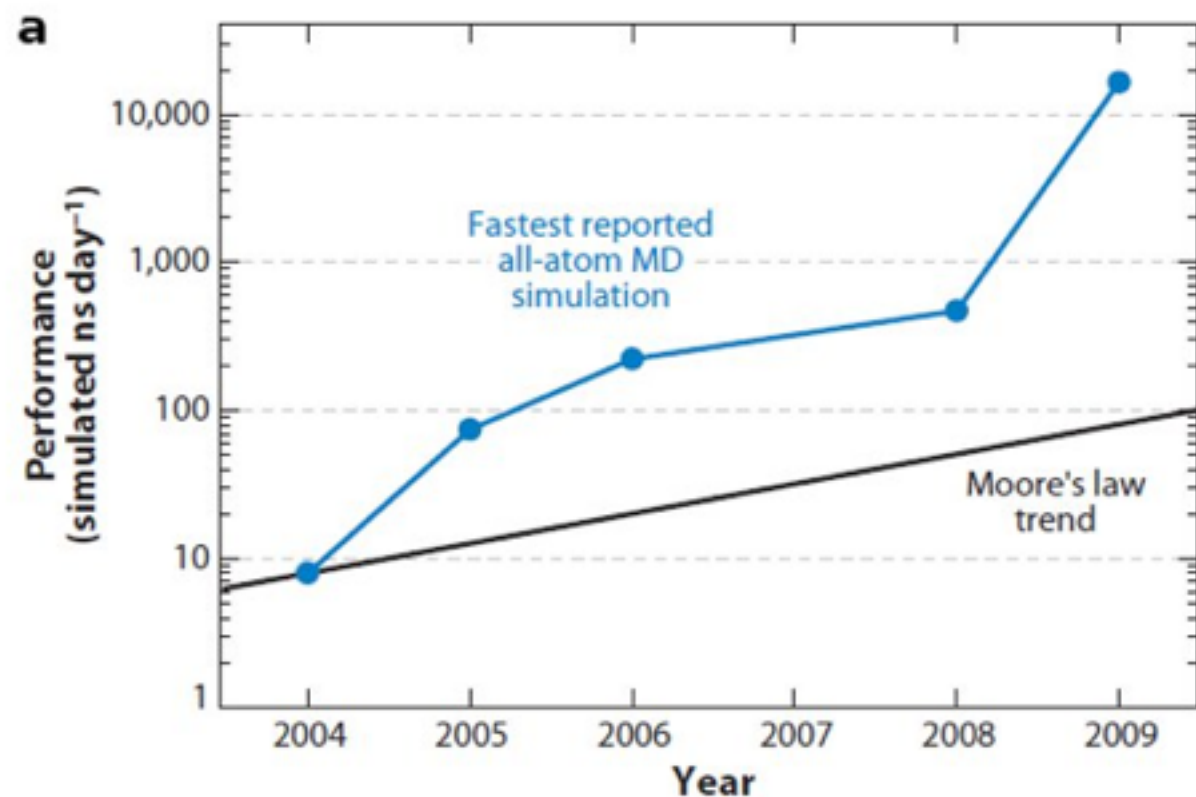
Pre-packaged tools and databases

- ▶ Many online
- ▶ New tools and time consuming methods frequently require downloading
- ▶ Most are free to use

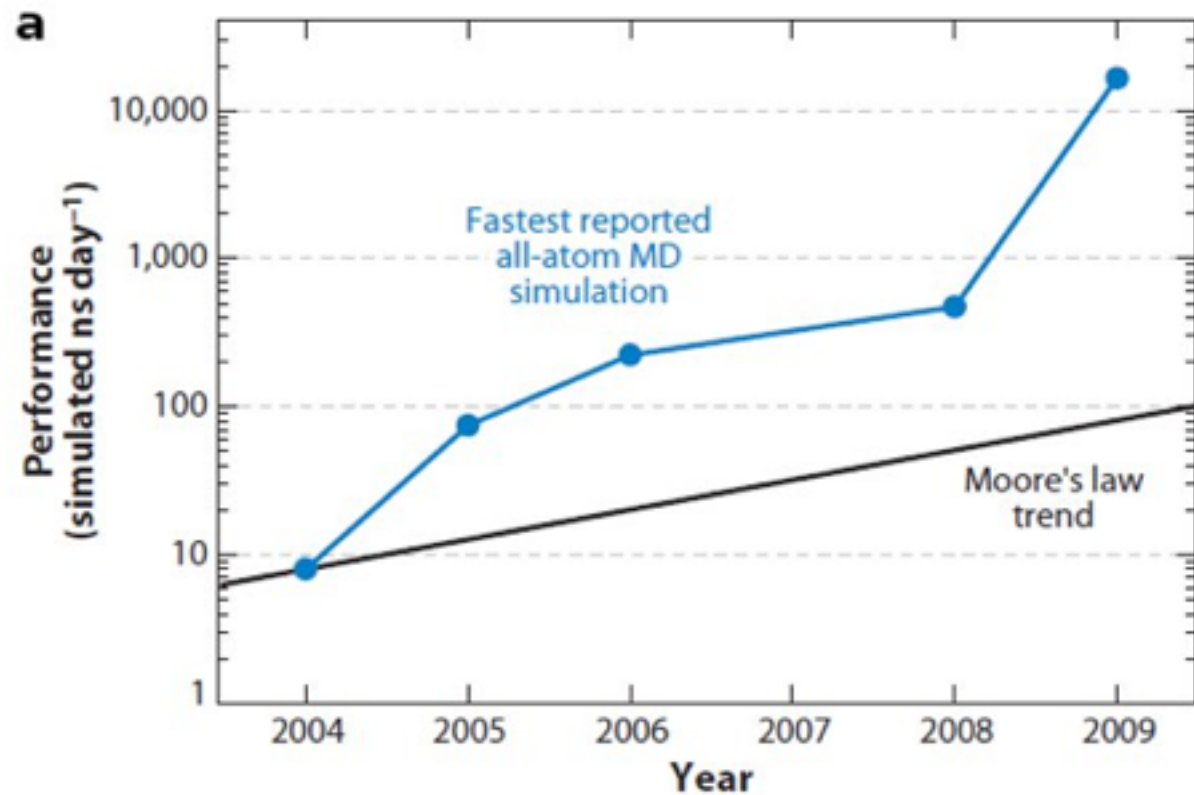
Tool development

- ▶ Mostly on a UNIX environment
- ▶ Knowledge of programming languages frequently required (Python, Perl, R, C Java, Fortran)
- ▶ May require specialized or high performance computing resources...

SIDE-NOTE: SUPERCOMPUTERS AND GPUS



SIDE-NOTE: SUPERCOMPUTERS AND GPUS



HOW COMPUTERS HAVE CHANGED

DATE	COST	SPEED	MEMORY	SIZE
1967	\$40M	0.1 MHz	1 MB	HALL
2013	\$4,000	1 GHz	10 GB	LAPTOP
CHANGE	10,000	10,000	10,000	10,000

If cars were like computers then a new Volvo would cost \$3, would have a top speed of 1,000,000 km/hr, would carry 50,000 adults and would park in a shoebox



Skepticism & Bioinformatics

We have to approach computational results the same way we do wet-lab results:

- Do they make sense?
- Is it what we expected?
- Do we have adequate controls, and how did they come out?
- Modeling is modeling, but biology is different...
What does this model actually contribute?
- Avoid the miss-use of 'black boxes'

Common problems with Bioinformatics

Confusing multitude of tools available

- ▶ Each with many options and settable parameters

Most tools and databases are written by and for nerds

- ▶ Same is true of documentation - if any exists!

Most are developed independently

Notable exceptions are found at the:

- **EBI** (European Bioinformatics Institute) and
- **NCBI** (National Center for Biotechnology Information)

General Parameters

Max target sequences	<input type="text" value="500"/>	Select the maximum number of aligned sequences to display ?
Short queries	<input checked="" type="checkbox"/> Automatically adjust parameters for short input sequences ?	
Expect threshold	<input type="text" value="10"/>	
Word size	<input type="text" value="3"/>	
Max matches in a query range	<input type="text" value="0"/>	

Scoring Parameters

Matrix	<input type="text" value="BLOSUM62"/>	
Gap Costs	<input type="text" value="Existence: 11 Extension: 1"/>	
Compositional adjustments	<input type="text" value="Conditional compositional scoring"/>	

Filters and Masking

Filter	<input type="checkbox"/> Low complexity regions ?	
Mask	<input type="checkbox"/> Mask for lookup table only ?	
	<input type="checkbox"/> Mask lower case letters ?	

PSI/PHI/DELTA BLAST

Upload PSSM	<input type="button" value="Choose File"/> no file selected	
PSI-BLAST Threshold	<input type="text" value="0.005"/>	
Pseudocount	<input type="text" value="0"/>	

Even Blast has many settable parameters

Related tools with different terminology

STEP 3 - Set your

PROGRAM

MATRIX	GAP OPEN	GAP EXTEND	KTUP	EXPECTATION UPPER VALUE	EXPECTATION LOWER VALUE
<input type="text" value="BLOSUM50"/>	<input type="text" value="-10"/>	<input type="text" value="-2"/>	<input type="text" value="2"/>	<input type="text" value="10"/>	<input type="text" value="0 (default)"/>
DNA STRAND	HISTOGRAM	FILTER	STATISTICAL ESTIMATES		
<input type="text" value="N/A"/>	<input type="text" value="no"/>	<input type="text" value="none"/>	<input type="text" value="Regress"/>		
SCORES	ALIGNMENTS	SEQUENCE RANGE	DATABASE RANGE	MULTI HSPs	
<input type="text" value="50"/>	<input type="text" value="50"/>	<input type="text" value="START-END"/>	<input type="text" value="START-END"/>	<input type="text" value="no"/>	
SCORE FORMAT					
<input type="text" value="Default"/>					

Key Online Bioinformatics Resources: NCBI & EBI

The NCBI and EBI are invaluable, publicly available resources for biomedical research

Please take the initial BIOINF525 questionnaire:
< tinyurl.com/bioinf525-questions >

<http://www.ncbi.nlm.nih.gov>

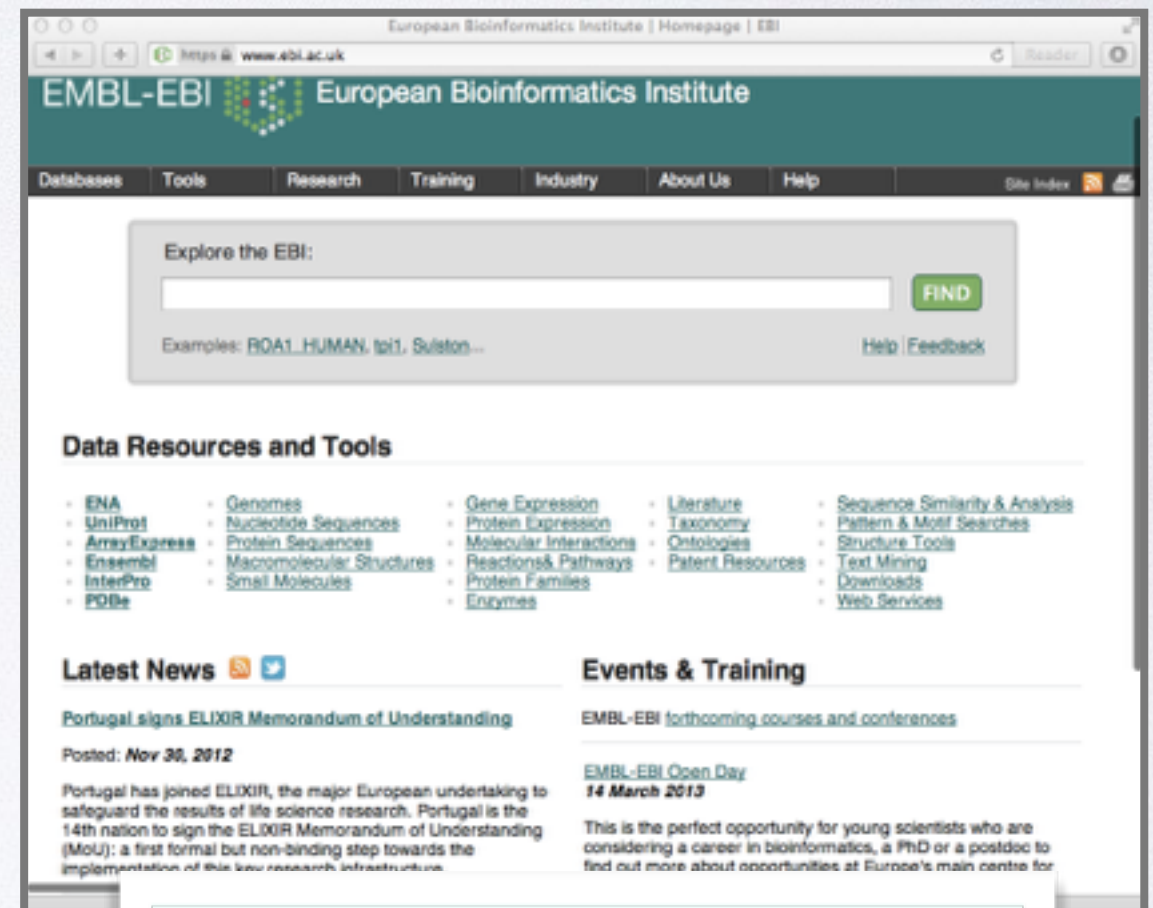
<https://www.ebi.ac.uk>

Key Online Bioinformatics Resources: NCBI & EBI

The NCBI and EBI are invaluable, publicly available resources for biomedical research



<http://www.ncbi.nlm.nih.gov>



<https://www.ebi.ac.uk>

National Center for Biotechnology Information (NCBI)

- Created in 1988 as a part of the National Library of Medicine (NLM) at the National Institutes of Health
- NCBI's mission includes:
 - ▶ Establish **public databases**
 - ▶ Develop **software tools**
 - ▶ **Education** on and dissemination of biomedical information
- We will cover a number of core NCBI databases and software tools in the lecture




<http://www.ncbi.nlm.nih.gov>

○ ○ ○

National Center for Biotechnology Information

◀ ▶ + 🔒 www.ncbi.nlm.nih.gov 🔍 Reader ⬇

NCBI Resources ▾ How To ▾ Sign in to NCBI



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.

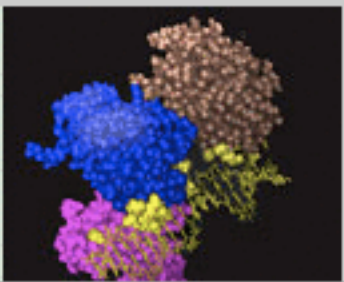
[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS Feeds](#)

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

3D Structures

Explore three-dimensional structures of proteins, DNA, and RNA molecules. Examine sequence-structure relationships, active sites, molecular interactions, biological activities of bound chemicals, and associated biosystems.



Popular Resources

PubMed

Bookshelf

PubMed Central

PubMed Health

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

NCBI Announcements

New version of Genome Workbench available

06 Sep

An integrated, downloadable applicati

<http://www.ncbi.nlm.nih.gov>

The image shows a screenshot of the National Center for Biotechnology Information (NCBI) website. A browser window is visible at the top with the address bar showing 'www.ncbi.nlm.nih.gov'. The NCBI logo and navigation links are at the top left. A search bar is at the top right. A 'Popular Resources' overlay box is positioned in the center-right, listing various resources with red arrows pointing to PubMed, BLAST, and a group of Nucleotide, Genome, SNP, Gene, Protein, and PubChem. The background website content includes a 'Welcome to NCBI' message, a 'Get Started' section with links to Tools, Downloads, How-To's, and Submissions, and a '3D Structures' section. A sidebar on the left lists various resource categories.

National Center for Biotechnology Information

www.ncbi.nlm.nih.gov

NCBI Resources How To

Sign in to NCBI

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 provides access to a wide range of biological and health information.

About the NCBI | Mission | Organization

Get Started

- Tools: Analyze data using NCBI tools
- Downloads: Get NCBI data
- How-To's: Learn how to access NCBI resources
- Submissions: Submit data to NCBI databases

3D Structures

Explore three-dimensional structures of proteins, DNA, and RNA molecules. Examine sequence-structure relationships, active sites, molecular interactions, biological activities of bound chemicals, and associated biosystems.

Popular Resources

PubMed

Bookshelf

PubMed Central

PubMed Health

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

Resources

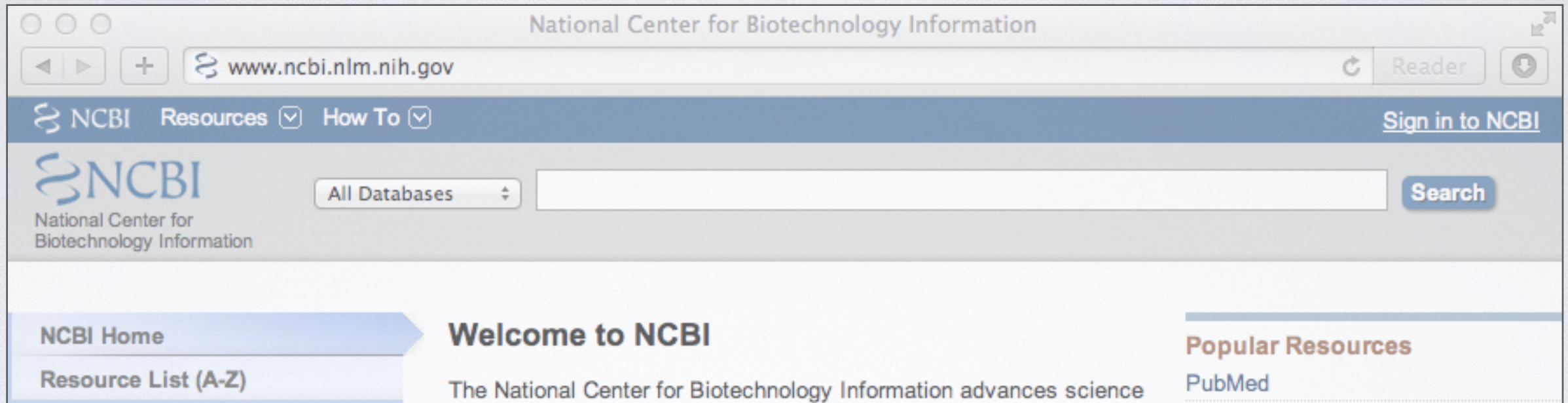
Announcements

New version of Genome Workbench available

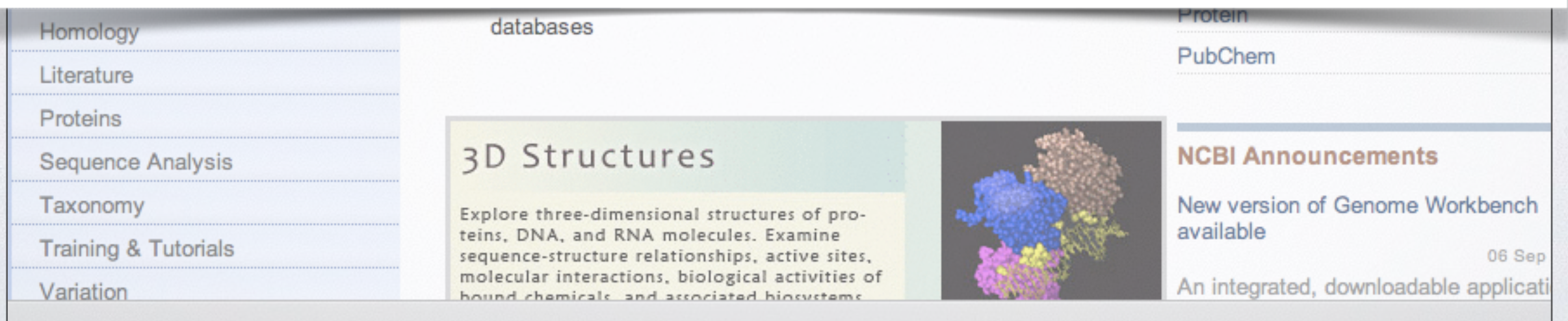
06 Sep

An integrated, downloadable application

<http://www.ncbi.nlm.nih.gov>

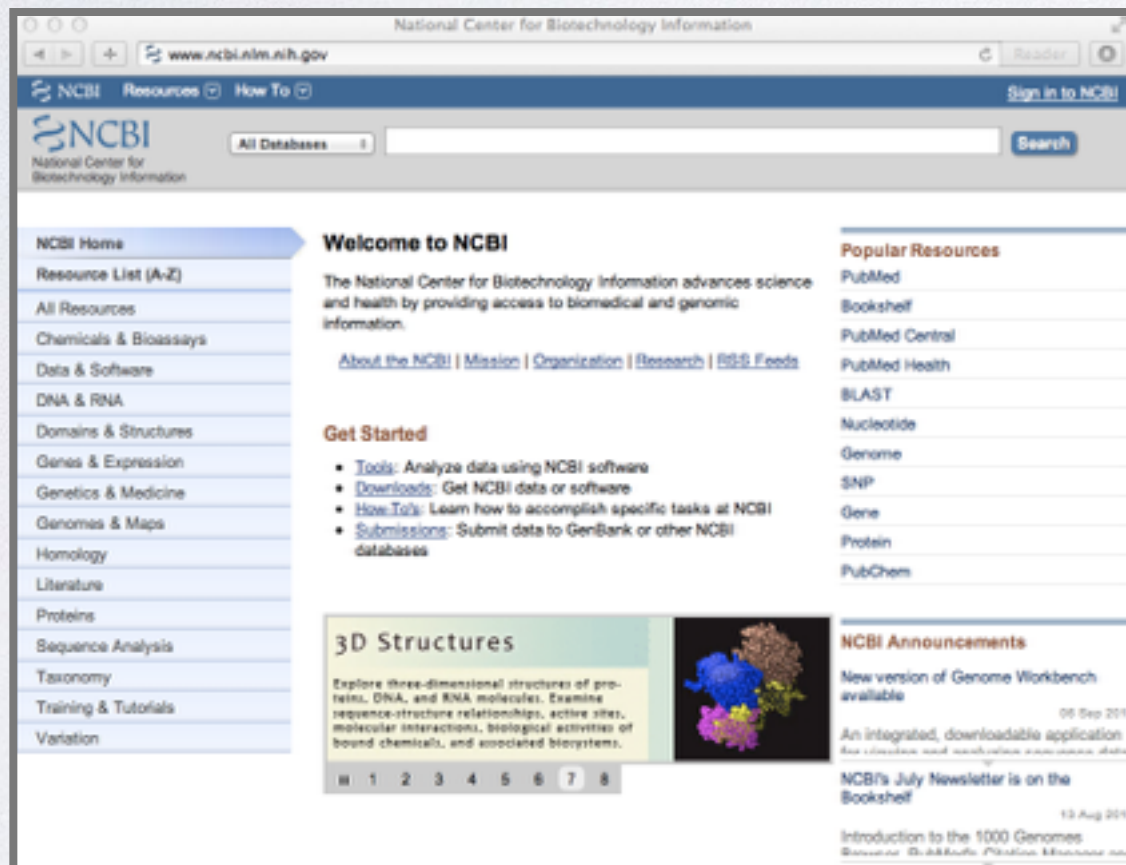


Notable NCBI databases include:
GenBank, RefSeq, PubMed, dbSNP
and the search tools ENTREZ and BLAST

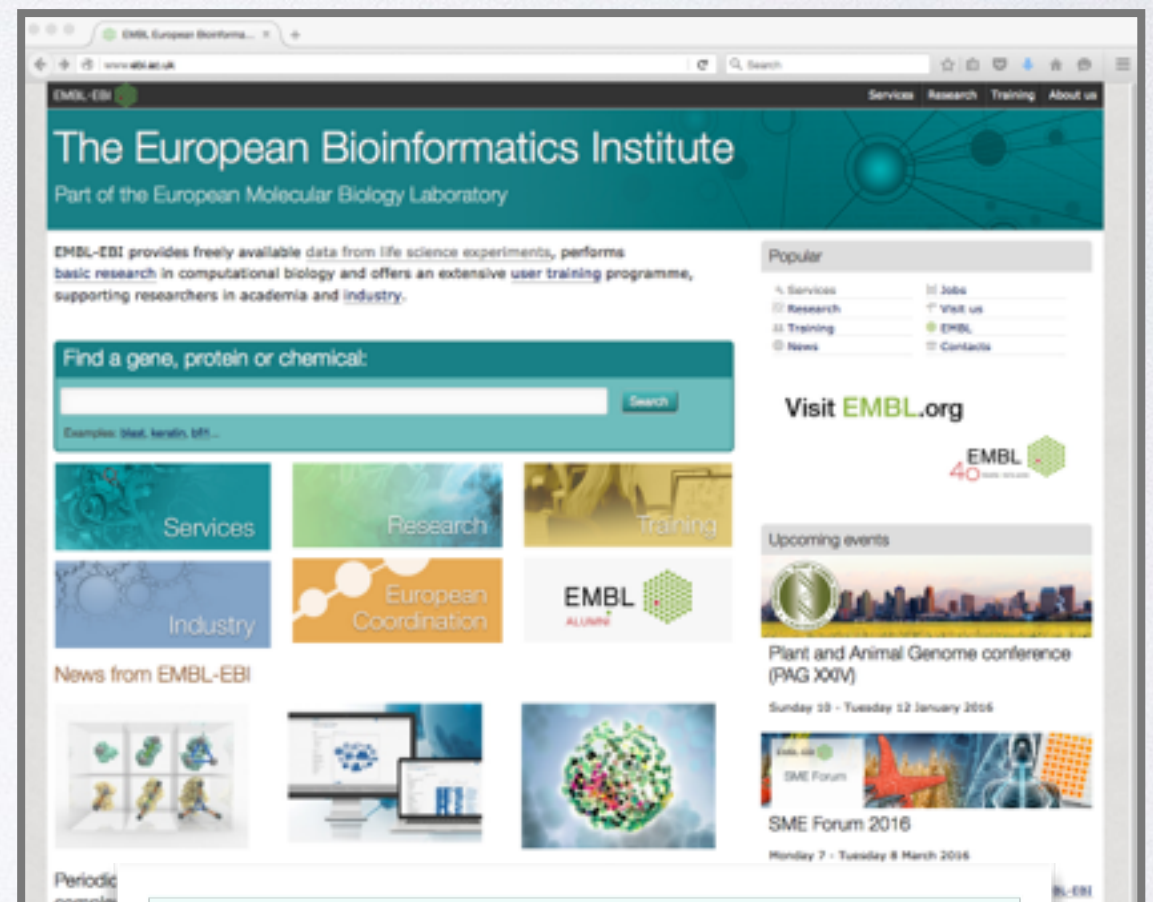


Key Online Bioinformatics Resources: NCBI & EBI

The NCBI and EBI are invaluable, publicly available resources for biomedical research



<http://www.ncbi.nlm.nih.gov>



<https://www.ebi.ac.uk>

European Bioinformatics Institute (EBI)

- Created in 1997 as a part of the European Molecular Biology Laboratory (EMBL)
- EBI's mission includes:
 - ▶ providing freely available **data and bioinformatics services**
 - ▶ and providing advanced **bioinformatics training**
- We will briefly cover several EBI databases and tools that have advantages over those offered at NCBI



The EBI maintains a number of high quality curated **secondary databases** and associated tools

The screenshot shows the EMBL-EBI website homepage. The browser address bar displays 'www.ebi.ac.uk'. The main header features the EMBL-EBI logo and navigation links for 'Services', 'Research', 'Training', and 'About us'. Below the header, a large teal banner reads 'The European Bioinformatics Institute' and 'Part of the European Molecular Biology Laboratory'. A paragraph states: 'EMBL-EBI provides freely available data from life science experiments, performs basic research in computational biology and offers an extensive user training programme, supporting researchers in academia and industry.' A search bar prompts users to 'Find a gene, protein or chemical:' with a 'Search' button and examples like 'blast', 'keratin', and 'bfl1...'. A grid of six colored tiles is visible: 'Services' (teal, highlighted with a red border), 'Research' (green), 'Training' (yellow), 'Industry' (blue), 'European Coordination' (orange), and 'EMBL ALUMNI' (white with a green hexagonal logo). To the right, a 'Popular' section lists links for Services, Research, Training, News, Jobs, Visit us, EMBL, and Contacts. Below this is a 'Visit EMBL.org' section with the EMBL 40th anniversary logo (1974-2014). The 'Upcoming events' section features a banner for the 'Plant and Animal Genome conference (PAG XXIV)' held from Sunday 10 to Tuesday 12 January 2016, accompanied by a city skyline image and a circular logo for the 'INTERNATIONAL PLANT & ANIMAL GENOME CONFERENCE'.

EMBL-EBI

Services Research Training About us

The European Bioinformatics Institute

Part of the European Molecular Biology Laboratory

EMBL-EBI provides freely available data from life science experiments, performs basic research in computational biology and offers an extensive user training programme, supporting researchers in academia and industry.

Find a gene, protein or chemical:

Search

Examples: blast, keratin, bfl1...

Services Research Training Industry European Coordination EMBL ALUMNI

Popular

- Services
- Research
- Training
- News
- Jobs
- Visit us
- EMBL
- Contacts

Visit **EMBL.org**

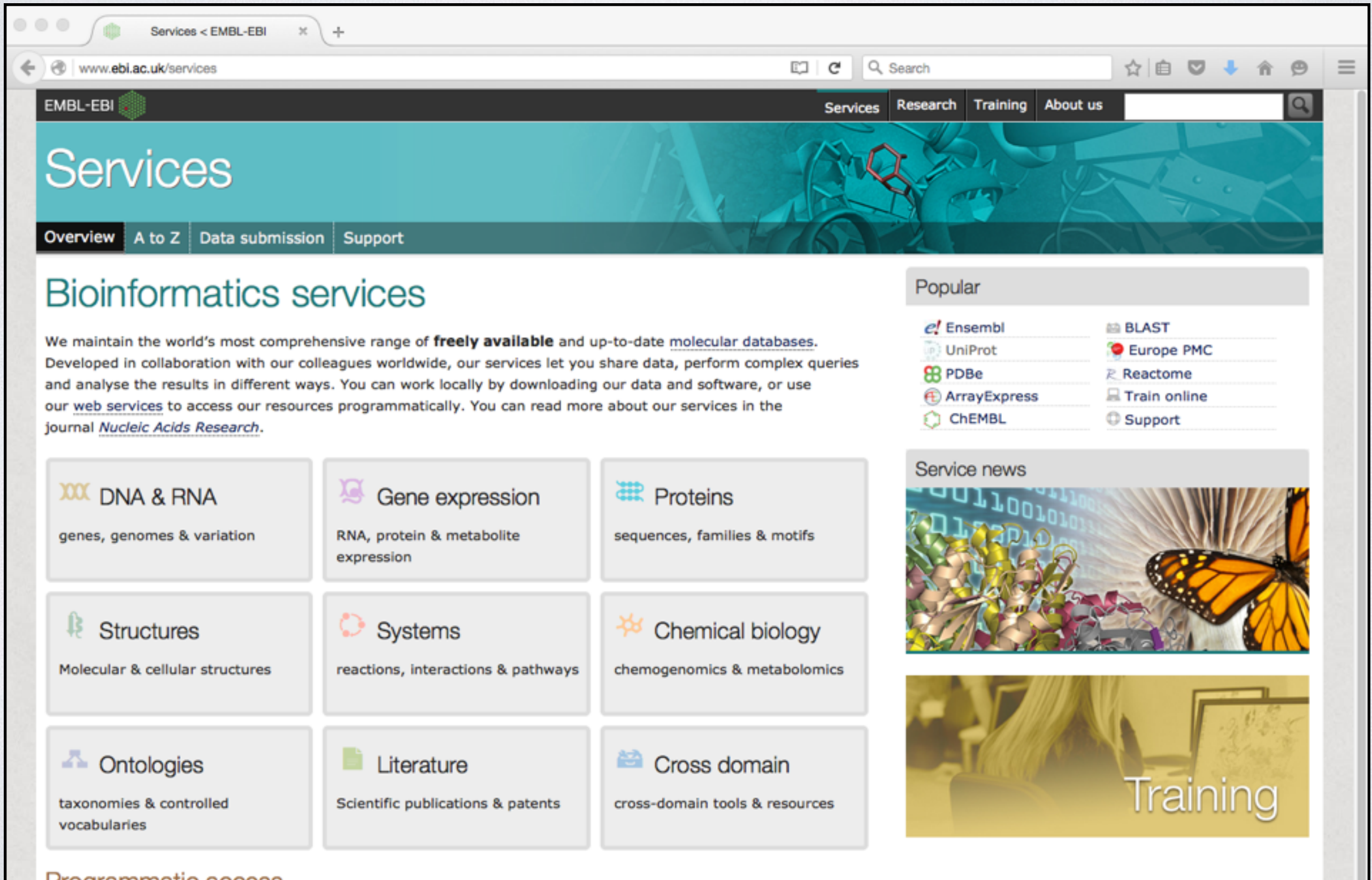
EMBL 40 YEARS 1974-2014

Upcoming events

Plant and Animal Genome conference (PAG XXIV)

Sunday 10 - Tuesday 12 January 2016

The EBI maintains a number of high quality curated **secondary databases** and associated tools



The screenshot shows the EMBL-EBI Services website. The browser address bar displays 'www.ebi.ac.uk/services'. The page features a teal header with the 'Services' title and navigation links for 'Overview', 'A to Z', 'Data submission', and 'Support'. A 'Bioinformatics services' section describes the availability of molecular databases. A 'Popular' sidebar lists tools like Ensembl, UniProt, PDB, and BLAST. The main content area is a grid of service categories: DNA & RNA, Gene expression, Proteins, Structures, Systems, Chemical biology, Ontologies, Literature, and Cross domain. A 'Service news' section with a butterfly image and a 'Training' section are also visible.

Services < EMBL-EBI

www.ebi.ac.uk/services

EMBL-EBI

Services Research Training About us

Services

Overview A to Z Data submission Support

Bioinformatics services

We maintain the world's most comprehensive range of **freely available** and up-to-date [molecular databases](#). Developed in collaboration with our colleagues worldwide, our services let you share data, perform complex queries and analyse the results in different ways. You can work locally by downloading our data and software, or use our [web services](#) to access our resources programmatically. You can read more about our services in the journal [Nucleic Acids Research](#).

DNA & RNA
genes, genomes & variation

Gene expression
RNA, protein & metabolite expression

Proteins
sequences, families & motifs

Structures
Molecular & cellular structures

Systems
reactions, interactions & pathways

Chemical biology
chemogenomics & metabolomics

Ontologies
taxonomies & controlled vocabularies

Literature
Scientific publications & patents

Cross domain
cross-domain tools & resources

Popular

- Ensembl
- UniProt
- PDB
- ArrayExpress
- ChEMBL
- BLAST
- Europe PMC
- Reactome
- Train online
- Support

Service news

Training

The EBI maintains a number of high quality curated **secondary databases** and associated tools

The screenshot shows the EMBL-EBI Services website. The browser address bar displays www.ebi.ac.uk/services. The page features a teal header with the 'Services' title and navigation links for 'Overview', 'A to Z', 'Data submission', and 'Support'. A main section titled 'Bioinformatics services' contains a paragraph about the availability of molecular databases. Below this is a grid of service categories, with 'Proteins' highlighted by a red box. To the right, a 'Popular' sidebar lists several databases, with 'Ensembl' and 'UniProt' also highlighted by a red box. At the bottom, there is a 'Training' section with a yellow background.

Services < EMBL-EBI

www.ebi.ac.uk/services

Services Research Training About us

Services

Overview A to Z Data submission Support

Bioinformatics services

We maintain the world's most comprehensive range of **freely available** and up-to-date molecular databases. Developed in collaboration with our colleagues worldwide, our services let you share data, perform complex queries and analyse the results in different ways. You can work locally by downloading our data and software, or use our web services to access our resources programmatically. You can read more about our services in the journal *Nucleic Acids Research*.

DNA & RNA
genes, genomes & variation

Gene expression
RNA, protein & metabolite expression

Proteins
sequences, families & motifs

Structures
Molecular & cellular structures

Systems
reactions, interactions & pathways

Chemical biology
chemogenomics & metabolomics

Ontologies
taxonomies & controlled vocabularies

Literature
Scientific publications & patents

Cross domain
cross-domain tools & resources

Popular

- Ensembl**
- UniProt**
- PDB**
- ArrayExpress**
- ChEMBL**








Training

<https://www.ebi.ac.uk>

The EBI makes available a wider variety of **online tools** than NCBI

Proteins

Popular services

	UniProt: The Universal Protein Resource The gold-standard, comprehensive resource for protein sequence and functional annotation data.
	InterPro A database for the classification of proteins into families, domains and conserved sites.
	PRIDE: The Proteomics Identifications Database An archive of protein expression data determined by mass spectrometry.
	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.
	HMMER - protein homology search Fast sensitive protein homology searches using profile hidden Markov models (HMMs). Variety of different search methods for querying against both sequence and HMM target databases.
	InterProScan 5 InterProScan 5 searches sequences against InterPro's predictive protein signatures. Please note that <u>InterProScan 4.8 has been retired</u> .

Quick links

- Popular services in this category
- All services in this category
- Project websites in this category

The EBI also provides a growing selection of **online tutorials** on EBI databases and tools

The screenshot shows the EMBL-EBI website homepage. The browser address bar displays 'www.ebi.ac.uk'. The main header features the EMBL-EBI logo and navigation links for 'Services', 'Research', 'Training', and 'About us'. The main banner reads 'The European Bioinformatics Institute' and 'Part of the European Molecular Biology Laboratory'. Below this, a paragraph states: 'EMBL-EBI provides freely available data from life science experiments, performs basic research in computational biology and offers an extensive user training programme, supporting researchers in academia and industry.' A search bar is present with the text 'Find a gene, protein or chemical:' and a 'Search' button. Below the search bar, examples are listed: 'blast, keratin, bfl1...'. The main content area is divided into several sections: 'Services', 'Research', 'Training' (highlighted with a red border), 'Industry', 'European Coordination', and 'EMBL ALUMNI'. On the right side, there is a 'Popular' section with links to 'Services', 'Research', 'Training', 'News', 'Jobs', 'Visit us', 'EMBL', and 'Contacts'. Below this is a 'Visit EMBL.org' section with the EMBL 40th anniversary logo (1974-2014). The 'Upcoming events' section features a banner for the 'Plant and Animal Genome conference (PAG XXIV)' held from Sunday 10 to Tuesday 12 January 2016. The bottom of the page shows a row of small images related to bioinformatics research.

EMBL-EBI

Services Research Training About us

The European Bioinformatics Institute

Part of the European Molecular Biology Laboratory

EMBL-EBI provides freely available data from life science experiments, performs basic research in computational biology and offers an extensive user training programme, supporting researchers in academia and industry.

Find a gene, protein or chemical:

Search

Examples: blast, keratin, bfl1...

Services Research Training Industry European Coordination EMBL ALUMNI

Popular

- Services
- Research
- Training
- News
- Jobs
- Visit us
- EMBL
- Contacts

Visit **EMBL.org**

EMBL 40 YEARS 1974-2014

Upcoming events

Plant and Animal Genome conference (PAG XXIV)

Sunday 10 - Tuesday 12 January 2016

The EBI also provides a growing selection of **online tutorials** on EBI databases and tools

The screenshot shows a web browser window with the URL www.ebi.ac.uk/training/online/course/using-sequence-similarity-searching-tools-embl-ebi. The page features a yellow header with the EMBL-EBI logo and navigation links: Services, Research, Training (selected), and About us. Below the header is a large banner with the text "Train online" and a background image of people working at computers. A secondary navigation bar includes links: Training, Train online Home, Course list, Glossary, Support & Feedback, and Log in / Register. The main content area is titled "Using sequence similarity searching tools at EMBL-EBI: webinar" and contains a video player. The video player shows a thumbnail with the text "Using sequence similarity search tools at EMBL-EBI" and "Finding homologous sequences with BLAST, FASTA, PSI-Search etc." along with a photo of Andrew Cowley and his contact information. The video player controls show a duration of 0:00 / 37:42. To the left of the video is a sidebar with "Course content" links: "Using sequence similarity searching tools at EMBL-EBI: webinar" (selected) and "Contributors", and a "Print Course" link. To the right is a "Popular" section with links: "Train online", "Find us", and "Funding", followed by a "Find us at..." section with links: "Open days and career days", "Conference exhibitions", "EMBL courses and events", "Genome campus events", and "Science for schools". At the bottom, a text box states: "This webinar focuses on how to use tools like **BLAST** and PSI-Search to find homologous sequences in EMBL-EBI databases, including tips on which tool and database to use, input formats, how to change parameters and how to interpret the results pages."

Using sequence similarity searching tools at EMBL-EBI: webinar

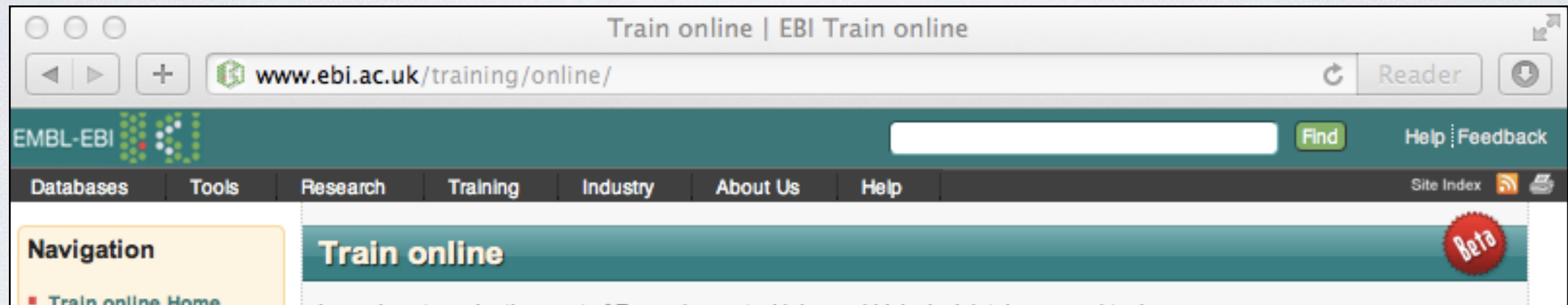
Using sequence similarity search tools at EMBL-EBI
Finding homologous sequences with BLAST, FASTA, PSI-Search etc.

Andrew Cowley
andrew.cowley@ebi.ac.uk
support@ebi.ac.uk

0:00 / 37:42

This webinar focuses on how to use tools like **BLAST** and PSI-Search to find homologous sequences in EMBL-EBI databases, including tips on which tool and database to use, input formats, how to change parameters and how to interpret the results pages.

The EBI also provides a growing selection of **online tutorials** on EBI databases and tools



Notable EBI databases include:

ENA, UniProt, Ensembl

and the tools FASTA, BLAST, InterProScan,
ClustalW, T-Coffee, MUSCLE, DALI, HMMER

Find a course

Browse by subject



[Genes and Genomes](#)



[Gene Expression](#)



[Interactions, Pathways and Networks](#)

BIOINFORMATICS DATABASES AND ASSOCIATED TOOLS

What is a database?

Computerized store of data that is organized to provide efficient retrieval.

- Uses standardized data (record) formats to enable computer handling

Key database features allow for:

- Adding, changing, removing and merging of records
- User-defined queries and extraction of specified records

Desirable features include:

- Contains the data you are interested in
- Allows fast data access
- Provides annotation and curation of entries
- Provides links to additional information (possibly in other databases)
- Allows you to make discoveries

Bioinformatics Databases

AATDB, AceDb, ACUTS, ADB, AFDB, AGIS, AMSdb, ARR, AsDb, BBDB, BCGD, Beanref, Biolmage, BioMagResBank, BIOMDB, BLOCKS, BovGBASE, BOVMAP, BSORF, BTKbase, CANSITE, CarbBank, CARBHYD, CATH, CAZY, CCDC, CD4OLbase, CGAP, ChickGBASE, Colibri, COPE, CottonDB, CSNDB, CUTG, CyanoBase, dbCFC, dbEST, dbSTS, DDBJ, DGP, DictyDb, Picty_cDB, DIP, DOGS, DOMO, DPD, DPlInteract, ECDC, ECGC, EC02DBASE, EcoCyc, EcoGene, EMBL, EMD db, ENZYME, EPD, EpoDB, ESTHER, FlyBase, FlyView, GCRDB, GDB, GENATLAS, Genbank, GeneCards, Genlilesne, GenLink, GENOTK, GenProtEC, GIFTS, GPCRDB, GRAP, GRBase, gRNAsdb, GRR, GSDB, HAEMB, HAMSTERS, HEART-2DPAGE, HEXAdb, HGMD, HIDB, HIDC, HIVdb, HotMolecBase, HOVERGEN, HPDB, HSC-2DPAGE, ICN, ICTVDB, IL2RGbase, IMGT, Kabat, KDNA, KEGG, Klotho, LGIC, MAD, MaizeDb, MDB, Medline, Mendel, MEROPS, MGDB, MGI, MHCPEP5 Micado, MitoDat, MITOMAP, MJDB, MmtDB, Mol-R-Us, MPDB, MRR, MutBase, MycDB, NDB, NRSub, O-lycBase, OMIA, OMIM, OPD, ORDB, OWL, PAHdb, PatBase, PDB, PDD, Pfam, PhosphoBase, PigBASE, PIR, PKR, PMD, PPDB, PRESAGE, PRINTS, ProDom, Prolysis, PROSITE, PROTOMAP, RatMAP, RDP, REBASE, RGP, SBASE, SCOP, SeqAnaiRef, SGD, SGP, SheepMap, Soybase, SPAD, SRNA db, SRPDB, STACK, StyGene, Sub2D, SubtiList, SWISS-2DPAGE, SWISS-3DIMAGE, SWISS- MODEL Repository, SWISS-PROT, TelDB, TGN, tmRDB, TOPS, TRANSFAC, TRR, UniGene, URNADB, V BASE, VDRR, VectorDB, WDCM, WIT, WormPep, etc !!!!

Bioinformatics Databases

There are lots of Bioinformatics Databases

For an annotated listing of major bioinformatics databases please see the [Ctools handout](#)

< [Handout_Major_Databases.pdf](#) >

Finding Bioinformatics Databases

The screenshot shows a web browser window with the URL www.oxfordjournals.org/nar/database/cat/8. The page is titled "Nucleic Acids Research" and features a navigation bar with links for "ABOUT THIS JOURNAL", "CONTACT THIS JOURNAL", "SUBSCRIPTIONS", "CURRENT ISSUE", "ARCHIVE", and "SEARCH". The main content area is titled "2014 NAR Database Summary Paper Category List" and lists various bioinformatics databases. A sidebar on the right contains a list of categories: "Compilation Paper", "Category List", "Alphabetical List", "Category/Paper List", and "Search Summary Papers".

Oxford Journals | Life Sciences | Nucleic Acids Research | Database Summary Paper Categories

www.oxfordjournals.org/nar/database/cat/8

Oxford Journals | Life Sciences | Nucleic Acids Research | Database Summary Pa... The 2014 Nucleic Acids Research Database Issue and an updated NAR online M...

OXFORD JOURNALS CONTACT US MY BASKET MY ACCOUNT

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals > Life Sciences > Nucleic Acids Research > Database Summary Paper Categories

2014 NAR Database Summary Paper Category List

Nucleotide Sequence Databases
RNA sequence databases
Protein sequence databases
Structure Databases
Genomics Databases (non-vertebrate)
Metabolic and Signaling Pathways
Human and other Vertebrate Genomes
Human Genes and Diseases
CancerResource
Protein Mutant Database
General human genetics databases
General polymorphism databases
Cancer gene databases
Gene-, system- or disease-specific databases
Microarray Data and other Gene Expression Databases
Proteomics Resources
Other Molecular Biology Databases
Organelle databases
Plant databases
Immunological databases
Cell biology

Compilation Paper
Category List
Alphabetical List
Category/Paper List
Search Summary Papers

<http://www.oxfordjournals.org/nar/database/c/>

Compilation Paper
Category List
Alphabetical List
Category/Paper List
Search Summary Papers

Oxford University Press is not responsible for the content of external internet sites

Major Molecular Databases

The most popular bioinformatics databases focus on:

- Biomolecular sequence (e.g. GenBank, UniProt)
- Biomolecular structure (e.g. PDB)
- Vertebrate genomes (e.g. Ensemble)
- Small molecules (e.g. PubChem)
- Biomedical literature (e.g. PubMed)

There are also many popular “*boutique*” databases for:

- Classifying protein families, domains and motifs (e.g. PFAM, PROSITE)
- Specific organisms (e.g. WormBase, FlyBase)
- Specific proteins of biomedical importance (e.g. KinaseDB, GPCRDB)
- Specific diseases, mutations (e.g. OMIM, HGMD)
- Specific fields or methods of study (e.g. GOA, IEDB)

Major Molecular Databases

The most popular bioinformatics databases focus on:

- Biomolecular sequence (e.g. GenBank, UniProt)
- Biomolecular structure (e.g. PDB)
- Vertebrate genomes (e.g. Ensemble)
- Small molecules (e.g. PubChem)
- Biomedical literature (e.g. PubMed)

There are also many "niche" databases for:

- Classifying proteins, domains and motifs (e.g. PFAM, PROSITE)
- Specific organisms (e.g. WormBase, FlyBase)
- Specific proteins of biomedical importance (e.g. KinaseDB, GPCRDB)
- Specific diseases, mutations (e.g. OMIM, HGMD)
- Specific fields or methods of study (e.g. GOA, IEDB)

See Online: [Handout Major Databases.pdf](#)

Primary, secondary & composite databases

Bioinformatics databases can be usefully classified into *primary*, *secondary* and *composite* according to their data source.

- **Primary databases** (or archival databases) consist of data derived experimentally.
 - ▶ **GenBank**: NCBI's primary nucleotide sequence database.
 - ▶ **PDB**: Protein X-ray crystal and NMR structures.
- **Secondary databases** (or derived databases) contain information derived from a primary database.
 - **RefSeq**: non redundant set of curated reference sequences primarily from GenBank
 - **PFAM**: protein sequence families primarily from UniProt and PDB
- **Composite databases** (or *metadatabases*) join a variety of different primary and secondary database sources.
 - **OMIM**: catalog of human genes, genetic disorders and related literature
 - **GENE**: molecular data and literature related to genes with extensive links to other databases.

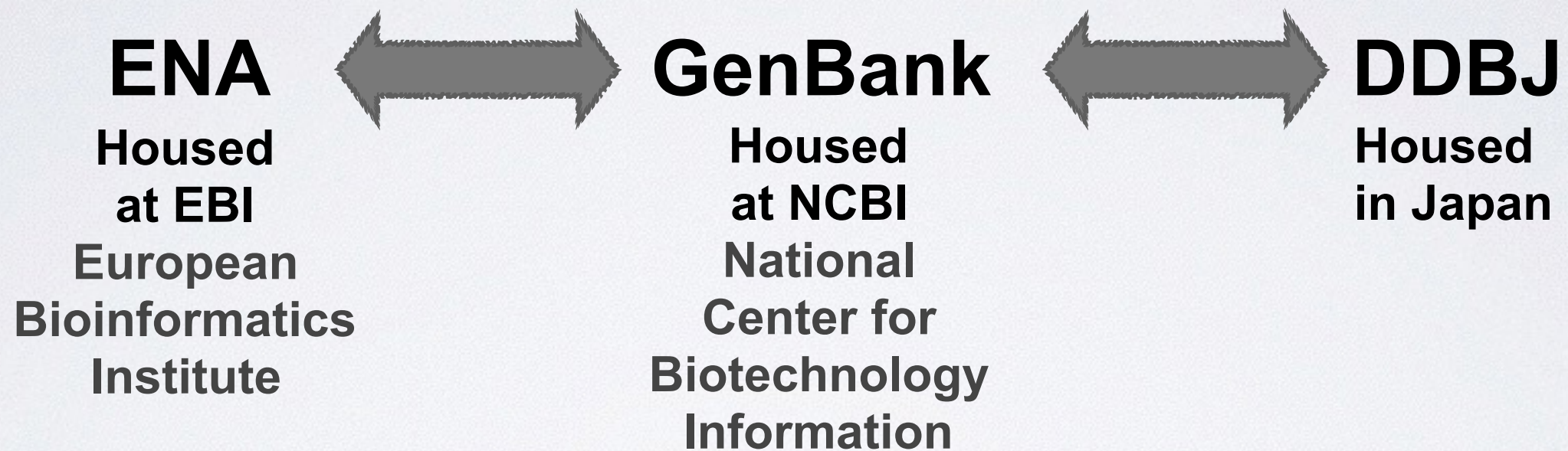
GENBANK & REFSEQ:
NCBI'S NUCLEOTIDE SEQUENCE
DATABASES

What is GenBank?

- GenBank is NCBI's primary **nucleotide only** sequence database
 - ▶ Archival in nature - reflects the state of knowledge at time of submission
 - ▶ Subjective - reflects the submitter point of view
 - ▶ Redundant - can have many copies of the same nucleotide sequence
- GenBank is actually three collaborating international databases from the US, Japan and Europe
 - ▶ GenBank (US)
 - ▶ DNA Database of Japan (DDBJ)
 - ▶ European Nucleotide Archive (ENA)

GenBank, ENA and DDBJ

Share and synchronize data



- The underlying raw DNA sequences are identical
 - ▶ The different sites provide different views and ways to navigate through the data
- Access to GenBank (and other NCBI databases including RefSeq) is typically through **Entrez**, (the Google of NCBI)
more on this later

GenBank sequence record

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

www.ncbi.nlm.nih.gov/nuccore/NM_004984.2

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide (KIF5A) AND "Homo sapiens" Search Limits Advanced Help

Display Settings: GenBank Send:

Homo sapiens kinesin family member 5A (KIF5A), mRNA

NCBI Reference Sequence: NM_004984.2

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

Go to:

LOCUS	NM_004984	3897 bp	mRNA	linear	PRI 10-JAN-2014
DEFINITION	Homo sapiens kinesin family member 5A (KIF5A), mRNA.				
ACCESSION	NM_004984				
VERSION	NM_004984.2	GI:45446748			
KEYWORDS	RefSeq.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 3897)				
AUTHORS	Kawaguchi, K.				
TITLE	Role of kinesin-1 in the pathogenesis of SPG10, a rare form of her				
JOURNAL	Neu				
PUBMED	227				
REMARK	Gen spa Rev				
REFERENCE	2				
AUTHORS	Pro Boh				
TITLE	alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay				
JOURNAL	J. Biol. Chem. 288 (30), 21742-21754 (2013)				
PUBMED	23744071				

GenBank flat file format has defined fields including unique identifiers such as the **ACCESSION** number.

This same general format is used for other sequence database records too.

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the KIF5A gene

alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay [J Biol Chem. 2013]

Peptide hormone metabolism

MHC class II antigen presentation

Side node: Database accession numbers

Database **accession numbers** are strings of letters and numbers used as **identifying labels** for sequences and other data within databases

► Examples (all for retinol-binding protein, RBP4):

X02775
NT_030059

GenBank genomic DNA sequence
Genomic contig

DNA

N91759.1
NM_006744

An expressed sequence tag (1 of 170)
RefSeq DNA sequence (from a transcript)

RNA

NP_007635
AAC02945
Q28369
1KT7

RefSeq protein
GenBank protein
UniProtKB/SwissProt protein
Protein Data Bank structure record

Protein

PMID: 12205585

PubMed IDs identify articles at NCBI/NIH

Literature

GenBank sequence record

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

www.ncbi.nlm.nih.gov/nuccore/NM_004984.2

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide (KIF5A) AND "Homo sapiens" Search Limits Advanced Help

Display Settings: GenBank Send:

Homo sapiens kinesin family member 5A (KIF5A), mRNA

NCBI Reference Sequence: NM_004984.2

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

Go to:

LOCUS	NM_004984	3897 bp	mRNA	linear	PRI 10-JAN-2014
DEFINITION	Homo sapiens kinesin family member 5A (KIF5A), mRNA.				
ACCESSION	NM_004984				
VERSION	NM_004984.2 GI:45446748				
KEYWORDS	RefSeq.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 3897)				
AUTHORS	Kawaguchi, K.				
TITLE	Role of kinesin-1 in the pathogenesis of SPG10, a rare form of hereditary spastic paraplegia				
JOURNAL	Neuroscientist 19 (4), 336-344 (2013)				
PUBMED	22785106				
REMARK	GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplegia type 10 when KIF5A is inactivated by mutations. Review article				
REFERENCE	2 (bases 1 to 3897)				
AUTHORS	Prots, I., Veber, V., Brey, S., Campioni, S., Buder, K., Riek, R., Bohm, K.J. and Winner, B.				
TITLE	alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay				
JOURNAL	J. Biol. Chem. 288 (30), 21742-21754 (2013)				
PUBMED	23744071				

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the KIF5A gene

[alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay \[J Biol Chem. 2013\]](#)

[Molecular motor KIF5A is essential for GABA\(A\) receptor transport, a \[Neuron. 2012\]](#)

[Systems-wide analysis of ubiquitylation dynamics reveals a key role \[Nat Cell Biol. 2012\]](#)

[See all...](#)

Pathways for the KIF5A gene

[Peptide hormone metabolism](#)

[MHC class II antigen presentation](#)

GenBank sequence record

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

www.ncbi.nlm.nih.gov/nuccore/NM_004984.2

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide (KIF5A) AND "Homo sapiens" Search Limits Advanced Help

Display Settings: GenBank Send: Change region shown

Homo sapiens kinesin family member 5A (KIF5A), mRNA

NCBI Reference Sequence: NM_004984.2

FASTA Graphics ← Can set different display formats here

Go to:

LOCUS	NM_004984	3897 bp	mRNA	linear	PRI 10-JAN-2014
DEFINITION	Homo sapiens kinesin family member 5A (KIF5A), mRNA.				
ACCESSION	NM_004984				
VERSION	NM_004984.2 GI:45446748				
KEYWORDS	RefSeq.				
SOURCE	Homo sapiens (human)				
ORGANISM	<u>Homo sapiens</u> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 3897)				
AUTHORS	Kawaguchi, K.				
TITLE	Role of kinesin-1 in the pathogenesis of SPG10, a rare form of hereditary spastic paraplegia				
JOURNAL	Neuroscientist 19 (4), 336-344 (2013)				
PUBMED	22785106				
REMARK	GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplegia type 10 when KIF5A is inactivated by mutations. Review article				
REFERENCE	2 (bases 1 to 3897)				
AUTHORS	Prots, I., Veber, V., Brey, S., Campioni, S., Buder, K., Riek, R., Bohm, K.J. and Winner, B.				
TITLE	alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay				
JOURNAL	J. Biol. Chem. 288 (30), 21742-21754 (2013)				
PUBMED	23744071				

Analyze this sequence

- Run BLAST
- Pick Primers
- Highlight Sequence Features
- Find in this Sequence

Articles about the KIF5A gene

- [alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay \[J Biol Chem. 2013\]](#)
- [Molecular motor KIF5A is essential for GABA\(A\) receptor transport, a \[Neuron. 2012\]](#)
- [Systems-wide analysis of ubiquitylation dynamics reveals a key role \[Nat Cell Biol. 2012\]](#)

See all...

Pathways for the KIF5A gene

- Peptide hormone metabolism
- MHC class II antigen presentation

FASTA sequence record

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

www.ncbi.nlm.nih.gov/nuccore/45446748?report=fasta

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Limits Advanced Help

Display Settings: FASTA Send: Change region shown Customize view

Homo sapiens kinesin family member 5A (KIF5A), mRNA

NCBI Reference Sequence: NM_004984.2

[GenBank](#) [Graphics](#)

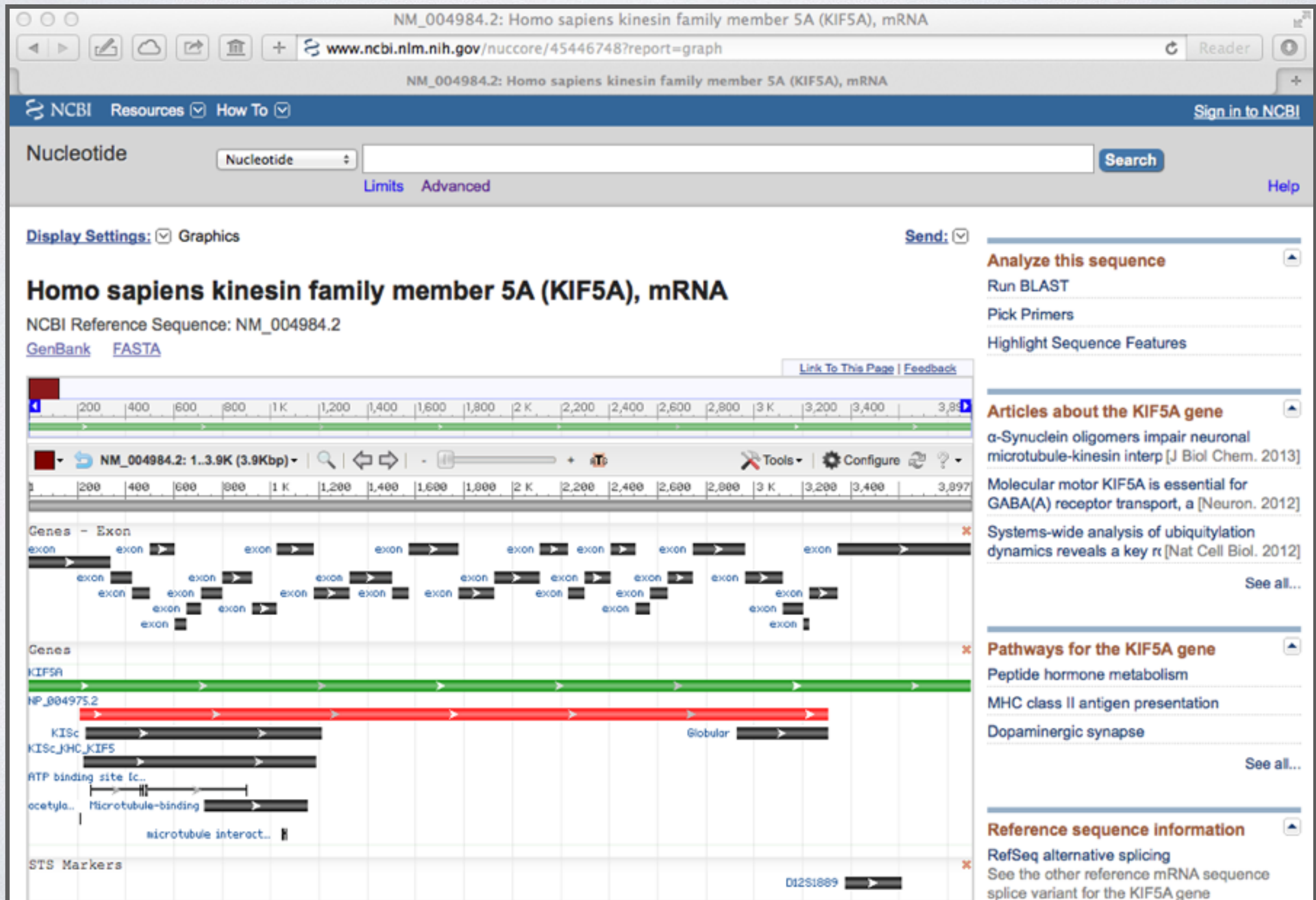
```
>gi|45446748|ref|NM_004984.2| Homo sapiens kinesin family member 5A (KIF5A), mRNA
ACGCCCAGGTGCGCCGTCATCCGCTGCCGAGGAGAGACAGCGCGCCCGGCCCTGCTCCCCAGGCTT
CGCCCGGGCGCCCTCAACTCTGTCCCAGAGACTGAGCACCTGTCTCCGCTCGGCCCTGTGCTGAGAGC
CCTCTCCTCTGGAGCACACACCACCCCTGCAGCCCAAGAAGAGTCCCAGCCCCACGCCGGCTACCACCAT
GGCGGAGACCAACAACGAATGTAGCATCAAGGTGCTCTGCCGATTCCGGCCCTGAACCAGGCTGAGATT
CTGCCGGGAGACAAGTTCATCCCCATTTTCCAAGGGGACGACAGCGTCTTATTGGGGGGAAGCCATATG
TTTTTGACCGTGTATTCCCCCAACACGACTCAAGAGCAAGTTTATCATGCATGTGCCATGCAGATTGT
CAAAGATGTCTTGTGCTGCTACAATGGCACCATTTTGTCTATGGACAGACATCCTCAGGGAAAACACAT
ACCATGGAGGGAAAGCTGCACGACCTCAGCTGATGGGAATCATTCCTCGAATTGCCCGAGACATCTTCA
ACCACATCTACTCCATGGATGAGAACCTTGAGTTCCACATCAAGGTTTCTTACTTTGAAATTTACCTGGA
CAAAATTCGTGACCTTCTGGATGTGACCAAGACAATCTGTCCGTGCACGAGGACAAGAACCGGGTGCCA
TTTGTCAAGGGTTGTACTGAACGCTTTGTGTCCAGCCCGGAGGAGATTCTGGATGTGATTGATGAAGGGA
AATCAAATCGTCATGTGGCTGTCAACAACATGAATGAACACAGCTCTCGGAGCCACAGCATCTTCCTCAT
CAACATCAAGCAGGAGAACATGGAACGGAGCAGAAGCTCAGTGGGAAGCTGTATCTGGTGGACCTGGCA
GGGAGTGAGAAGGTCAGCAAGACTGGAGCAGAGGGAGCCGTGCTGGACGAGGCAAGAATATCAACAAGT
CACTGTGAGCTCTGGGCAATGTGATCTCCGCACTGGCTGAGGGCACTAAAAGCTATGTTCCATATCGTGA
CAGCAAAATGACAAGGATTCTCCAGGACTCTCTCGGGGAAACTGCCGGACGACTATGTTTCATCTGTTGC
TCACCATCCAGTTATAATGATGCAGAGACCAAGTCCACCCTGATGTTTGGGCAGCGGGCAAAGACCATTA
AGAACACTGCCTCAGTAAATTTGGAGTTGACTGCTGAGCAGTGGAAGAAGAAATATGAGAAGGAGAAGGA
GAAGACAAAGGGCCAGAGGAGACGATTGCGAAGCTGGAGGCTGAGCTGAGCCGGTGGCGCAATGGAGAG
AATGTGCTGAGACAGAGCGCTGGCTGGGGAGGAGGAGCCCTGGGAGCCGAGCTCTGTGAGGAGACCC
CTGTGAATGACAACCTCATCCATCGTGGTGCGCATCGCGCCCGAGGAGCGGCAGAAATACGAGGAGGAGAT
CCGCCGTCTCTATAAGCAGCTTGACGACAAGGATGATGAAATCAACCAACAAAGCCAACTCATAGAGAAG
CTCAAGCAGCAAAATGCTGGACCAGGAAGAGCTGCTGGTGTCCACCCGAGGAGACAACGAGAAGGTCCAGC
GGGAGCTGAGCCACCTGCAATCAGAGAACGATGCCGCTAAGGATGAGGTGAAGGAAGTGTGTCAGGCCCT
GGAGGAGCTGGCTGTGAACTATGACCAGAAGTCCCAGGAGGTGGAGGAGAAGAGCCAGCAGAACCAGCTT
CTGGTGGATGAGCTGTCTCAGAAGGTGGCCACCATGCTGTCCCTGGAGTCTGAGTTGCAGCGGCTACAGG
AGGTCAGTGGACACCAGCGAAAACGAATTGCTGAGGTGCTGAACGGGCTGATGAAGGATCTGAGCGAGTT
```

FASTA sequence files consist of records where each record begins with a “>” and header information on that same line. Each subsequent line of the record is sequence information.

This format is commonly used by sequence analysis programs.

Pathways for the KIF5A gene
Peptide hormone metabolism
MHC class II antigen presentation

GenBank 'graphics' sequence record



GenBank sequence record, cont.

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

www.ncbi.nlm.nih.gov/nuccore/NM_004984.2

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide (KIF5A) AND "Homo sapiens" Search Limits Advanced Help

Display Settings: GenBank Send:

Homo sapiens kinesin family member 5A (KIF5A), mRNA

NCBI Reference Sequence: NM_004984.2

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

Go to:

LOCUS	NM_004984	3897 bp	mRNA	linear	PRI 10-JAN-2014
DEFINITION	Homo sapiens kinesin family member 5A (KIF5A), mRNA.				
ACCESSION	NM_004984				
VERSION	NM_004984.2 GI:45446748				
KEYWORDS	RefSeq.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 3897)				
AUTHORS	Kawaguchi, K.				
TITLE	Role of kinesin-1 in the pathogenesis of SPG10, a rare form of hereditary spastic paraplegia				
JOURNAL	Neuroscientist 19 (4), 336-344 (2013)				
PUBMED	22785106				
REMARK	GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplegia type 10 when KIF5A is inactivated by mutations. Review article				
REFERENCE	2 (bases 1 to 3897)				
AUTHORS	Prots, I., Veber, V., Brey, S., Campioni, S., Buder, K., Riek, R., Bohm, K.J. and Winner, B.				
TITLE	alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay				
JOURNAL	J. Biol. Chem. 288 (30), 21742-21754 (2013)				
PUBMED	23744071				

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the KIF5A gene

[alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay \[J Biol Chem. 2013\]](#)

[Molecular motor KIF5A is essential for GABA\(A\) receptor transport, a \[Neuron. 2012\]](#)

[Systems-wide analysis of ubiquitylation dynamics reveals a key role \[Nat Cell Biol. 2012\]](#)

[See all...](#)

Pathways for the KIF5A gene

[Peptide hormone metabolism](#)

[MHC class II antigen presentation](#)

GenBank sequence record, cont.

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

www.ncbi.nlm.nih.gov/nuccore/45446748?report=genbank&to=3897#feature_45446748

Reader

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

FEATURES	Location/Qualifiers	OMIM
source	1..3897 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /chromosome="12" /map="12q13.13"	Probe
gene	1..3897 /gene="KIF5A" /gene_synonym="D12S1889; MY050; NKHC; SPG10" /note="kinesin family member 5A" /db_xref="GeneID:3798" /db_xref="HGNC:6323" /db_xref="HPRD:09108" /db_xref="MIM:602821"	Protein PubMed PubMed (RefSeq)
exon	1..337 /gene="KIF5A" /gene_synonym="D12S1889; MY050; NKHC; SPG10" /inference="alignment:Splign:1.39.8"	
misc_feature	134..136 /gene="KIF5A" /gene_synonym="D12S1889; MY050; NKHC; SPG10" /note="upstream in-frame stop codon"	
CDS	209..3307 /gene="KIF5A" /gene_synonym="D12S1889; MY050; NKHC; SPG10" /note="kinesin, heavy chain, neuron-specific; KIF5A variant protein; neuronal kinesin heavy chain; kinesin heavy chain neuron-specific 1" /codon_start=1 /product="kinesin heavy chain isoform 5A" /protein_id="NP_004975.2" /db_xref="GI:45446749" /db_xref="CCDS:CCDS8945.1" /db_xref="GeneID:3798" /db_xref="HGNC:6323" /db_xref="HPRD:09108" /db_xref="MIM:602821" /translation="MAETNNECSIKVLCRFRPLNQAEILRGDKFIPIFQGDDSVVIGG KPYVFDVRVFPNTTQEQVYHACAMQIVKDVLAGYNGTIFAYGQTSSGKTHMEGKLHD PQLMGIIIPRIARDIFNHIYSMDENLEPHIKVSYPFIYLDKIRDLLDVTKTNLSVHEDK NRVPFVKGCTERFVSSPEEILDVIDEGKSNRHVAVTNMNEHSSRSHSIFLINIKQENM ETEQLSGKLYLVDLAGSEKVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKSY VPYRDSKMTRILQDSLGCNCRTTTFICCPSSYNDAETKSTLMFCQRAKTIKNTASVN"	

The **FEATURES** section contains annotations including a conceptual translation of the nucleotide sequence.

Recent activity

Turn Off Clear

Homo sapiens kinesin family member 5A (KIF5A), mRNA Nucleotide

(kinesin) AND "Homo sapiens"[porgn] (1351) Nucleotide

kinesin (37064) Nucleotide

See more...

GenBank sequence record, cont.

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

www.ncbi.nlm.nih.gov/nuccore/45446748?report=genbank&to=3897#sequence_45446748

Reader

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

/gene="KIF5A"
/gene_synonym="D12S1889; MY050; NKHC; SPG10"
/standard_name="D12S1889"
/db_xref="UniSTS:48006"

ORIGIN

```
1  acgcccaggt  cgcccgcatc  ccgctgccc  aggagagaga  cagcgcgccc  cggccctgct
61  ccccaggttt  cgcccgggcg  cctcactc  tgtcccaga  gactgagcac  ctgtcctccg
121  cctcggtctc  tcttgagagc  cctctcct  ggagcacaca  ccaccctgc  agcccaagaa
181  ggtgcccagc  cccacgccc  ctaccacct  ggaggagacc  aacaacgaat  gtagcatcaa
241  ggtgctctgc  cgattccggc  ccctgaacca  ggctgagatt  ctgcggggag  acaagttcat
301  cccatttttc  caaggggacg  acagcgtcgt  tattgggggg  aagccatatg  tttttgaccg
361  tgtattcccc  ccaaacacga  ctcaagagca  agtttatcat  gcatgtgcca  tgcagattgt
421  caaagatgtc  cttgctggct  acaatggcac  catttttgct  tatggacaga  catcctcagg
481  gaaaacacat  accatggagg  gaaagctgca  cgaccctcag  ctgatgggaa  tcattcctcg
541  aattgcccga  gacatcttca  accacatcta  ctccatggat  gagaaccttg  agttccacat
601  caaggtttct  tactttgaaa  tttacctgga  caaaattcgt  gaccttctgg  atgtgaccaa
661  gacaaatctg  tccgtgcacg  aggacaagaa  ccgggtgcca  tttgtcaagg  gttgtactga
721  acgctttgtg  tccagcccg  aggagattct  ggatgtgatt  gatgaaggga  aatcaaatcg
781  tcatgtggtc  gtcaccaaca  tgaatgaaca  cagctctcgg  agccacagca  tcttcctcat
841  caacatcaag  caggagaaca  tggaaacgga  gcagaagctc  agtgggaagc  tgtatctggt
901  ggacctggca  gggagtgaga  aggtcagcaa  gactggagca  gagggagccg  tgctggacga
961  ggcaaagaat  atcaacaagt  cactgtcagc  tctgggcaat  gtgatctccg  cactggctga
1021  gggcactaaa  agctatgttc  catatcgtga  cagcaaatg  acaaggattc  tccaggactc
1081  tctcggggga  aactgccgga  cgactatgtt  catctgttgc  tcaccatcca  gttataatga
1141  tgcagagacc  aagtccacce  tgatgtttgg  gcagcgggca  aagaccatta  agaacactgc
1201  ctcatgaaat  ttggagtga  ctgctgagca  gtggaagaag  aaatatgaga  aggagaagga
1261  gaagacaaag  gcccgaaagg  agacgattgc  gaagctggag  gctgagctga  gccgggtggc
1321  caatggagag  aatgtgcctg  agacagagcg  cctggctggg  gaggaggcag  ccctgggagc
1381  cgagctctgt  gaggagaccc  ctgtgaatga  caactcatcc  atcgtggtgc  gcatcgcgcc
1441  cgaggagcgg  cagaaatacg  aggaggagat  ccgcccgtct  tataagcagc  ttgacgacaa
1501  ggatgatgaa  atcaaccaac  aaagccaact  catagagaag  ctcaagcagc  aaatgctgga
1561  ccaggaagag  ctgctggtgt  ccaccgagg  agacaacgag  aaggtccagc  gggagctgag
1621  ccacctgcaa  tcagagaacg  atgccgctaa  ggatgaggtg  aaggaaagtgc  tgcaggccct
1681  ggaggagctg  gctgtgaact  atgaccagaa  gtcccaggag  gtggaggaga  agagccagca
1741  gaaccagctt  ctggtggatg  agctgtctca  gaaggtggcc  accatgctgt  ccctggagtc
1801  tgagttgcag  cggctacagg  aggtcagtg  acaccagcga  aaacgaattg  ctgaggtgct
1861  gaacgggctg  atgaaggatc  tgagcaggtt  cagtgtcatt  gtgggcaacg  gggagattaa
1921  gctgccagt  gagatcagtg  gggccatcga  ggaggagtgc  actgtggccc  gactctacat
1981  cagcaaaatc  aaatcagaag  tcaagtctgt  ggtcaagcgg  tgccggcagc  tggagaacct
2041  ccaggtggag  tgtaccgca  agatggaagt  gaccggcgcg  gagctctcat  cctgccagct
2101  cctcatctct  cagcatgagg  ccaagatccg  ctgccttacg  gaatacatgc  agagcgtgga
2161  gctaaagaag  cggcacctgg  aagagtcta  tgactccttg  agcagtgagc  tggccaagct
2221  ccaggcccag  gaaactgtgc  atgaagtggc  cctgaaggac  aaggagcctg  acactcagga
2281  tgcagatgaa  gtgaagaagg  ctctggagct  gcagatggag  agtcaccggg  aggcccatca
2341  ccggcagctg  gcccggtccc  gggacgagat  caacgagaag  cagaagacca  ttgatgagct
```

The actual sequence entry starts after the word **ORIGIN**

RefSeq: NCBI's Derivative Sequence Database

- RefSeq entries are hand curated best representation of a transcript or protein (in their judgement)
- Non-redundant for a given species although alternate transcript forms will be included if there is good evidence

- Experimentally verified transcripts and proteins
accession numbers begin with “NM_” or “NP_”
- Model transcripts and proteins based on bioinformatics predictions with little experimental support
accession numbers begin with “XM_” or “XP_”
- RefSeq also contains contigs and chromosome records

UNIPROT:
THE PREMIER PROTEIN SEQUENCE
DATABASE

UniProt: Protein sequence database

UniProt is a comprehensive, high-quality resource of protein sequence and functional information

- UniProt comprises four databases:

1. **UniProtKB** (Knowledgebase)

Containing **Swiss-Prot** and **TrEMBL** components
(these correspond to hand curated and automatically annotated entries respectively)

2. **UniRef** (Reference Clusters)

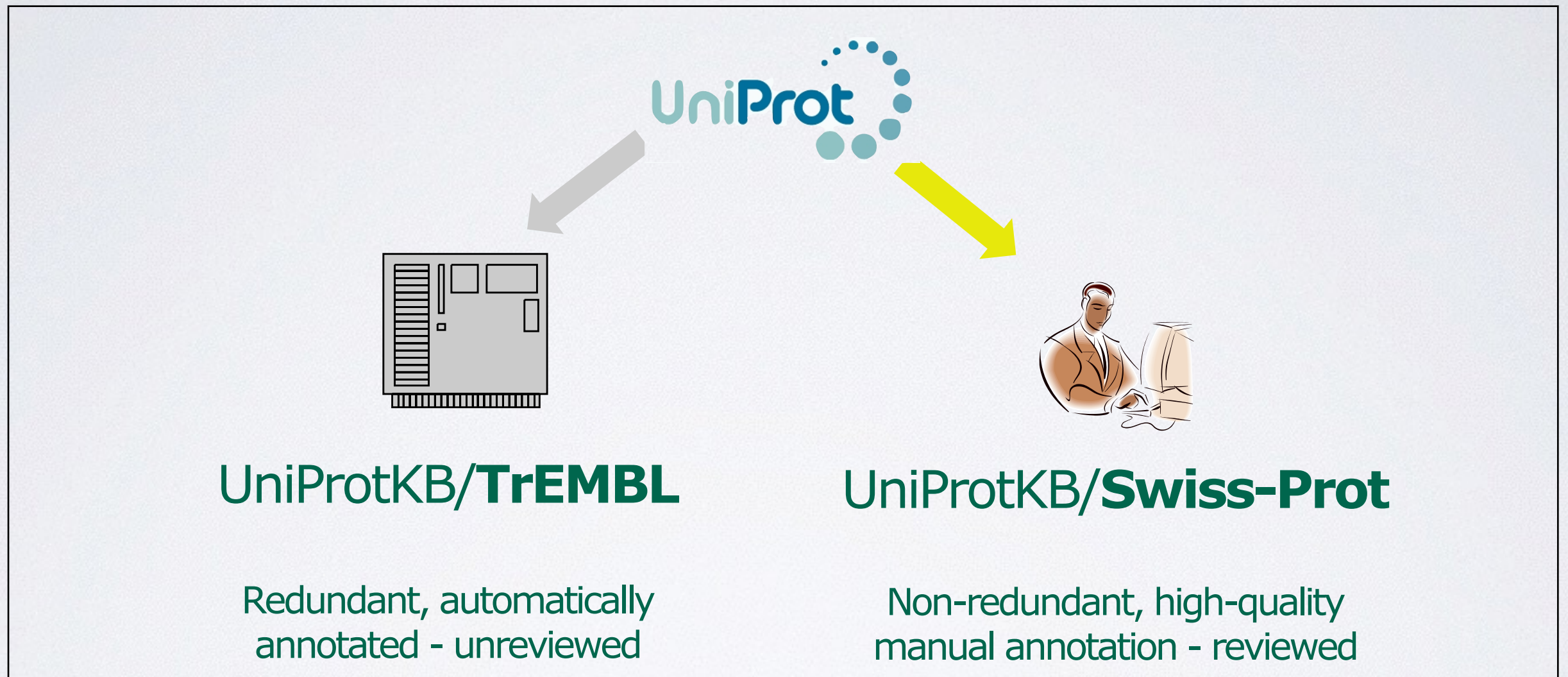
Filtered version of UniProtKB at various levels of sequence identity

e.g. UniRef90 contains sequences with a maximum of 90% sequence identity to each other

3. **UniParc** (Archive) with database cross-references to source.

4. **UniMES** (Metagenomic and Environmental Sequences)

The two sides of UniProtKB



★ Unreviewed, UniProtKB/TrEMBL **Q9N0H9** (Q9N0H9_EQUAS)

★ Reviewed, UniProtKB/Swiss-Prot **P38398** (BRCA1_HUMAN)

Indicators of which part of UniProt an entry belongs to include the color of the stars and the ID

The main information added to a UniProt/Swiss-Prot entry

- [1] "The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins."
Ebersole T.A., Chen Q., Justice M.J., Artzt K.
Nat. Genet. 12:260-265(1996) [PubMed: 8589716] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 3), INVOLVEMENT IN QKV, TISSUE SPECIFICITY, MUTAGENESIS OF
- [2] "Genomic organization and expression of the quaking gene."
Kondo T., Furuta T., Mitsuhashi K., Ebersole T.A., Shichiri M., Wu J., Artzt K., Yamamura K., Abe K.
Mamm. Genome 10:662-669(1999) [PubMed: 10384037] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA] (ISOFORMS 2; 3; 4 AND 7), ALTERNATIVE SPLICING (ISOFORM 1).
Strain: 129/J.

General annotation (Comments)		Hide	Top
Function	RNA-binding protein that plays a central role in myelination. Also required for visceral endoderm function and blood vessel development. Binds to the 5'-NACUAAU-N(120)-UAAU-3' RNA core sequence. Acts by regulating pre-mRNA splicing, mRNA export, mRNA stability and protein translation, as well as cellular processes including apoptosis, cell cycle, glial cell fate and development. Required to protect and promote stability of mRNAs such as MBP and CDKN1B to promote oligodendrocyte differentiation. Participates in mRNA transport by regulating the nuclear export of MBP mRNA. Isoform 1 is involved in regulation of mRNA splicing of MAG pre-mRNA by acting as a negative regulator of MAG exon 12 alternative splicing. Isoform 3 can induce apoptosis, while heterodimerization with other isoforms result in nuclear translocation of isoform 3 and suppression of apoptosis. Isoform 4 acts as a translational repressor for GLUT1. May also play a role in smooth muscle development.		
Subunit structure	Homodimer. Does not require RNA to homodimerize. Able to heterodimerize with BICC1.		
Subcellular location	Cytoplasm; Nucleus. Note-Isoform 1 localizes predominantly in the nucleus and at lower level in cytoplasm. It shuttles between the cytoplasm and the nucleus. Isoform 3 localizes predominantly in the cytoplasm and at much lower level in nucleus. Isoform 4 localizes both in the cytoplasm and nucleus.		
Tissue specificity	Highly expressed in neuro-forming cells. Expressed in oligodendrocytes and astrocytes in the central nervous system as well as Schwann cells. Isoform 1 is expressed in the mesodermal site of developing blood vessels. Isoform 3 is expressed in the brain, lung, heart and testis.		
Developmental stage	Neuronal differentiation. By contrast, neural progenitors located in specific sub-domains of the vz maintain expression as they differentiate and migrate away into the emerging nervous system. These have characteristics consistent with the acquisition of a glial rather than neuronal fate (at protein level). First detected in the neuroepithelium of mouse head folds at E7.5. Expression is strongly present ventrally in the recent brain and neural tube of E8.5 and E9.5 and in the heart. Isoform 1 is expressed in early embryos, while isoform 3 and isoform 4 are expressed in late embryos.		
Post-translational modification	Tyrosine phosphorylation and serine phosphorylation are involved in the regulation of protein function. Phosphorylation is an important activity, affecting transport and/or stabilization of MBP mRNA. The level of Tyr phosphorylation in the developing myelin is highest in the first postnatal week (P7). During the vigorous accumulation of MBP mRNA between P7 and P20, phosphorylation in the developing myelin drastically declined. By the end of the fourth postnatal week (P28), phosphorylation is reduced approximately 90%.		
Involvement in disease	Defects in Qkv are the cause of quaking/shake (qkv). Qkv is a spontaneous mutation resulting in hypomyelination of the central and peripheral nervous systems. Mutant mice develop normally until postnatal day 10 when they display rapid tremors or 'quaking' that is especially pronounced in hindlimbs and experience convulsive tonic-clonic seizures as they mature. Mice with qkv specifically lack isoform 3.		

Cell cycle	Regulation of cell proliferation
DNA damage	Traceable author statement. Source: UniProtKB
DNA repair	Regulation of transcription from RNA polymerase II promoter
Fatty acid biosynthesis	Traceable author statement. Source: Protinc
Lipid synthesis	Regulation of transcription from RNA polymerase III promoter
Nucleus	Traceable author statement. Source: UniProtKB
Polymorphism	Response to estrogen stimulus
Disease mutation	BRCA1-BARD1 complex
Repeat	Inferred from direct assay. Source: UniProtKB
Zinc-finger	Gamma-tubulin ring complex
DNA-binding	Non-traceable author statement. Source: UniProtKB
Metal-binding	DNA binding
Zinc	Traceable author statement. Source: Protinc
Anti-oncogene	Androgen receptor binding
Phosphorylation	Non-traceable author statement. Source: UniProtKB
3D-structure	Enzyme binding
	Inferred from physical interaction. Source: UniProtKB

```

      10      20      30      40      50      60
MVGEMETKEK PKPTPDYIMQ LMNDKKLMSS LPNFCGIFNH LERLLDEEIS RVRKDMYNDT

      70      80      90     100     110     120
LNGSTEKRSÄ ELPDVAGPIV QLQEKLYVPV KEYPDFNFVG RILGPRGLTÄ KQLEAETGCK

     130     140     150     160     170     180
IMVVGKGSMT DYQKQNDG KKNWEHLNED LHVLIIVVEDÄ ONPÄEIKLKR AVEEVKKLLV

     190     200     210     220     230     240
PAAEGEDSLK KMQLMELAIL NGTLDANIK SPALAFSLAA TAQAAPRIIT GPAPVLPPAA

     250     260     270     280     290     300
LRTPTPAGPT IMPLIRIQIT AVMPNGTPHP TAAIVPPGPE AGLIYTPYEV PYTLAPATSI

     310     320     330     340
LEYPIEPSGV LGAVATKVRÄ HDMRVHPYQR IVTADRAATG N
```

Isoform 1 (identifier: Q9QYS9-1)
Also known as QK1-5;
This isoform has been chosen as the 'canonical' sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

Isoform 2 (identifier: Q9QYS9-2)
Also known as QK1-7b;
The sequence of this isoform differs from the canonical sequence as follows:
312-341 GAVATKVRHDMRVHPYQRIVTADRAATGN → VWLSQRKAKNSRTVLTEPSSDLNLNTNA

Isoform 3 (identifier: Q9QYS9-3)
Also known as Qk1-7;
The sequence of this isoform differs from the canonical sequence as follows:
312-341 GAVATKVRHDMRVHPYQRIVTADRAATGN → EWIEMPVMPDISAH

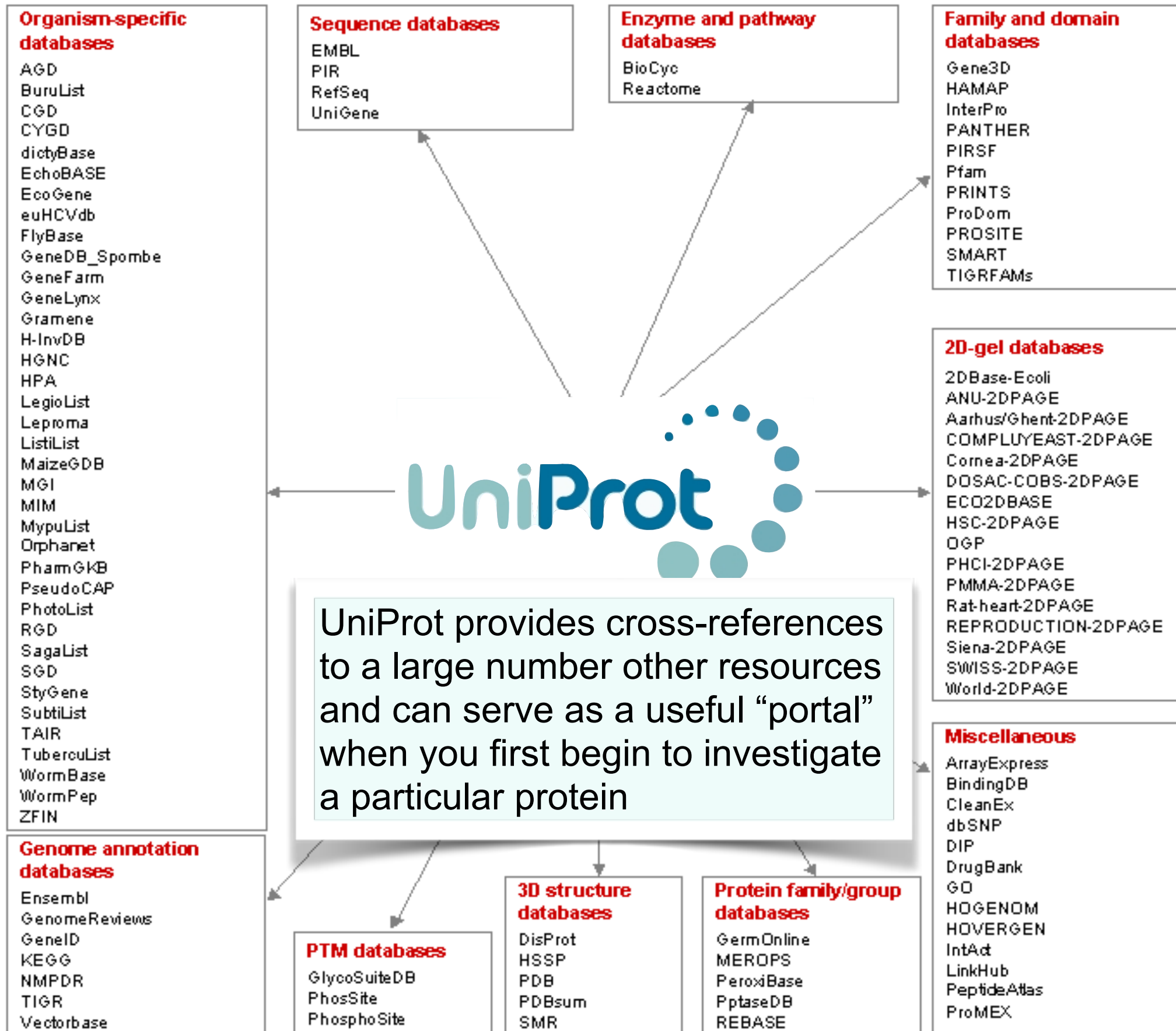
Protein quaking
Also known as:
Mqkl

Nomenclature

Gene names	Name: Qk1
	Synonyms: Qk, Qk1, Qka1

Molecule processing				
<input type="checkbox"/> Chain	1 – 341	341	Protein quaking	<div></div>
Regions				
<input type="checkbox"/> Domain	87 – 153	67	KH	<div></div>
<input type="checkbox"/> Motif	276 – 311	36	Protein quaking	<div></div>
<input type="checkbox"/> Motif	324 – 330	7	Nuclear localization signal	<div></div>

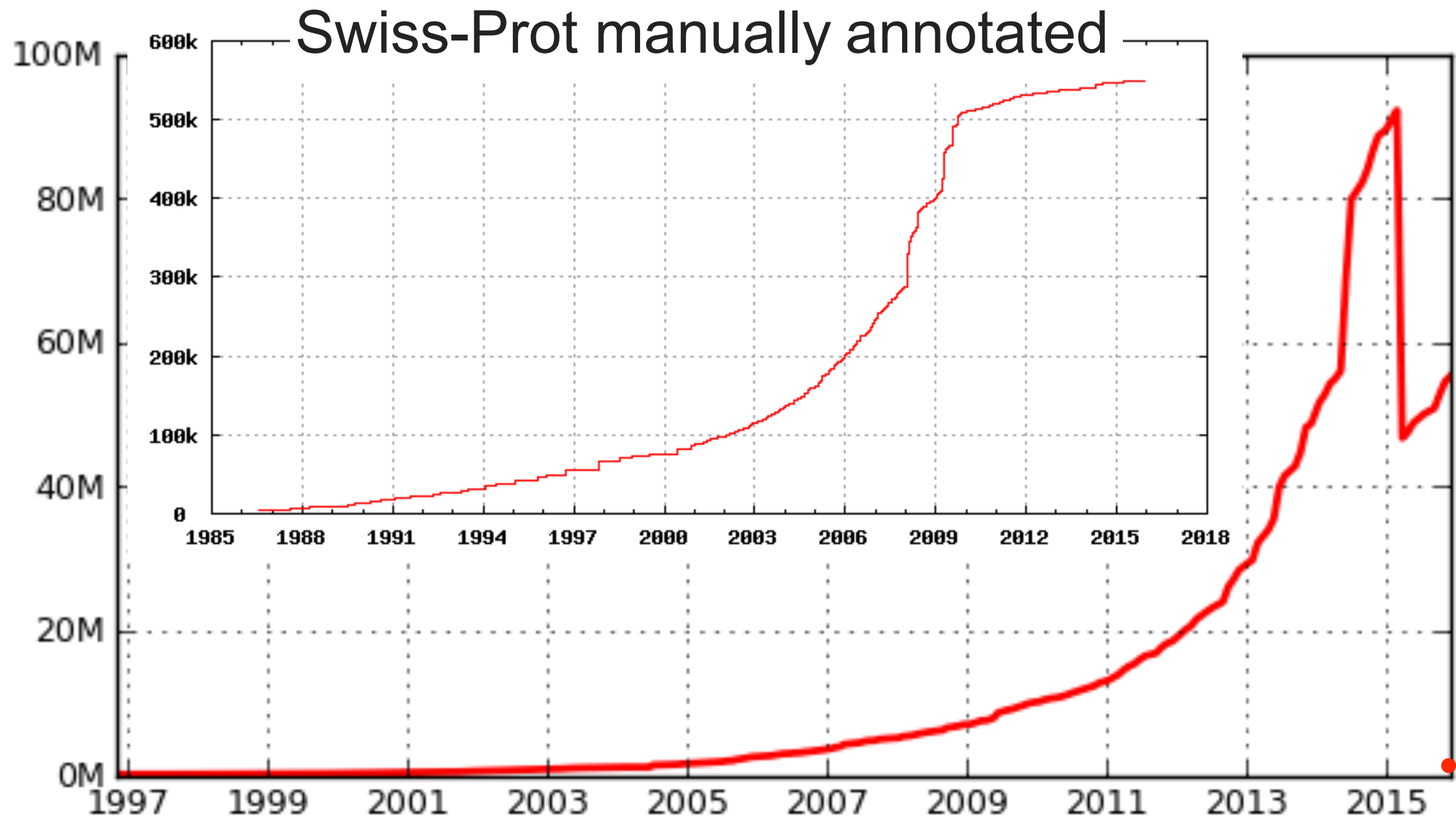
UniProt / Swiss-Prot



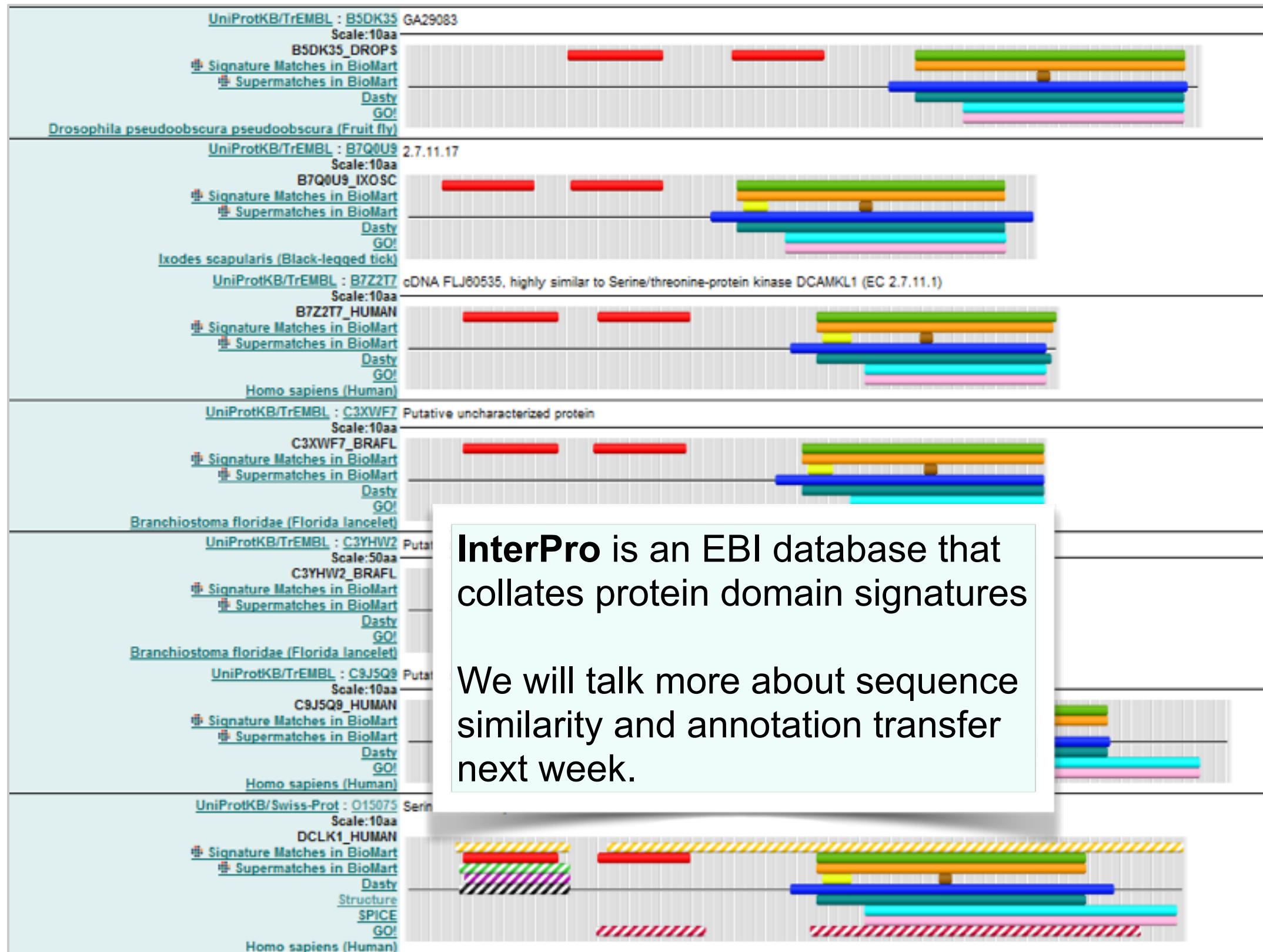
UniProt/Swiss-Prot vs UniProt/TrEMBL

- UniProtKB/Swiss-Prot is a **non-redundant** database with one entry per protein
- UniProtKB/TrEMBL is a **redundant** database with one entry per translated ENA entry (ENA is the EBI's equivalent of GenBank)
 - ▶ Therefore TrEMBL can contain multiple entries for the same protein
 - ▶ Multiple UniProtKB/TrEMBL entries for the same protein can arise due to:
 - Erroneous gene model predictions
 - Sequence errors (Frame shifts)
 - Polymorphisms
 - Alternative start sites
 - Isoforms
 - OR because the same sequence was submitted by different people

Side note: Automatic Annotation (sharing the wealth)



Same domain composition = same function = annotation transfer



InterPro is an EBI database that collates protein domain signatures

We will talk more about sequence similarity and annotation transfer next week.



DATABASE VIGNETTE

You have just come out a seminar about gastric cancer and one of your co-workers asks:

“What do you know about that ‘Kras’ gene the speaker kept taking about?”

You have some recollection about hearing of ‘Ras’ before. How would you find out more?

- Google?
- Library?
- **Bioinformatics databases at NCBI and EBI!**

<http://www.ncbi.nlm.nih.gov/>

<http://www.ncbi.nlm.nih.gov/>

The image is a screenshot of the National Center for Biotechnology Information (NCBI) website. The browser's address bar shows the URL www.ncbi.nlm.nih.gov/. The NCBI logo and navigation links like 'Resources' and 'How To' are visible at the top. A search bar is present with the text 'All Databases' and a dropdown menu showing 'ras', which is highlighted with a red square. To the right of the search bar is a 'Search' button. On the left side, there is a vertical menu with various resource categories such as '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', and 'Variation'. The main content area features a 'Welcome to NCBI' message, a description of the center's mission, and links to 'About the NCBI', 'Mission', 'Organizations', and 'NCBI News'. Below this, there is a 'Get Started' section with links to 'Data using NCBI software', 'Get NCBI data or software', 'Learn how to accomplish specific tasks at NCBI', and 'Submit data to GenBank or other NCBI databases'. A large, white, diagonal banner with red text reading 'Hands on demo (or see following slides)' is overlaid across the center of the page. At the bottom, there is a section titled 'Genotypes and Phenotypes' with text about genome-wide association studies, and a sidebar on the right with 'NCBI Announcements' regarding RefSeq release 69.

NCBI Resources How To Sign in to NCBI

All Databases ras 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 and health by providing access to biomedical information.

About the NCBI | Mission | Organizations | NCBI News

Get Started

- [Data using NCBI software](#)
- [Get NCBI data or software](#)
- [Learn how to accomplish specific tasks at NCBI](#)
- [Submissions](#): Submit data to GenBank or other NCBI databases

Genotypes and Phenotypes

Data from Genome Wide Association studies that link genes and diseases. See study variables, protocols, and analysis.

NCBI Announcements

RefSeq release 69 available on

The full RefSeq release 69 is not available on the FTP site with 74 records describing 52,278,488

Search NCBI databases

Help

ras

Search

About 2,978,774 search results for "ras"

Literature

Books	1,677	books and reports
MeSH	402	ontology used for PubMed indexing
NLM Catalog	223	books, journals and more in the NLM Collections
PubMed	54,672	scientific & medical abstracts/citations
PubMed Central	96,114	full-text journal articles

Health

ClinVar	759	human variations of clinical significance
dbGaP	120	genotype/phenotype interaction studies
GTR	1,879	genetic testing registry

Genes

EST	3,985	expressed sequence tag sequences
Gene	87,165	collected information about gene loci
GEO DataSets	3,732	functional genomics studies
GEO Profiles	1,622,789	gene expression and molecular abundance profiles
HomoloGene	696	homologous gene sets for selected organisms
PopSet	2,254	sequence sets from phylogenetic and population studies
UniGene	4,770	clusters of expressed transcripts

Proteins

Gene Gene ras Search
 Save search Advanced Help

Show additional filters

Clear all

Gene sources

Genomic
 Mitochondria
 Organelles
 Plasmids
 Plastids

Categories

Alternatively spliced
 Annotated genes
 Non-coding
 Protein-coding
 Pseudogene

Sequence

content
 CCDS
 Ensembl
 RefSeq

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

Send to:

Hide sidebar >>

Did you mean ras as a gene symbol?
 Search Gene for ras as a symbol.

<< First < Prev Page 1 of 4282 Next > Last >>

Results: 1 to 20 of 85633

Filters activated: Current only. Clear all to show 87165 items.

Filters: Manage Filters

Top Organisms [Tree]

Homo sapiens (1126)
 Mus musculus (823)
 Rattus norvegicus (625)
 Oreochromis niloticus (533)
 Neolamprologus brichardi (507)
 All other taxa (82019)

More...

Find related data

Database:

Select

Find items

Search details

ras[All Fields] AND
 alive[property]

Name/Gene ID	Description	Location	Aliases
<input type="checkbox"/> ras ID: 19412	resistance to audiogenic seizures [<i>Mus musculus</i> (house mouse)]		asr
<input type="checkbox"/> ras ID: 43873	rasberry [<i>Drosophila melanogaster</i> (fruit fly)]	Chromosome X, NC_004354.4 (10744502..10749097)	Dmel_CG1799, CG11485, CG1799, Dmel\CG1799, EP(X)1093,

Gene

Search

[Help](#)

Display Settings: ☒ Tabular, 20 per page, Sorted by Relevance

Send to:

Hide sidebar >>

Filters: [Manage Filters](#)

Results: 1 to 20 of 1126 << First < Prev Page 1 of 57 Next > Last >>

 Filters activated: Current only. [Clear all](#) to show 1499 items.

Name/Gene ID	Description	Location	Aliases
<input type="checkbox"/> NRAS ID: 4893	neuroblastoma RAS viral (v- ras) oncogene homolog [<i>Homo sapiens</i> (human)]	Chromosome 1, NC_000001.11 (114704464..114716894, complement)	RP5- 1000E10.2, ALPS4, CMNS, N-ras, NCMS1, NS6, NRAS
<input type="checkbox"/> KRAS ID: 3845	Kirsten rat sarcoma viral oncogene homolog [<i>Homo sapiens</i> (human)]	Chromosome 12, NC_000012.12 (25205246..25250923, complement)	C-K-RAS, CFC2, K- RAS2A, K- RAS2B, K- RAS4A, K- RAS4B, KI- RAS1, KRAS2, NS, NS3, RASK2

Find related data

Database:

Select

Search details

```
ras[All Fields] AND "Homo  
sapiens"[porgn] AND  
alive[property]
```

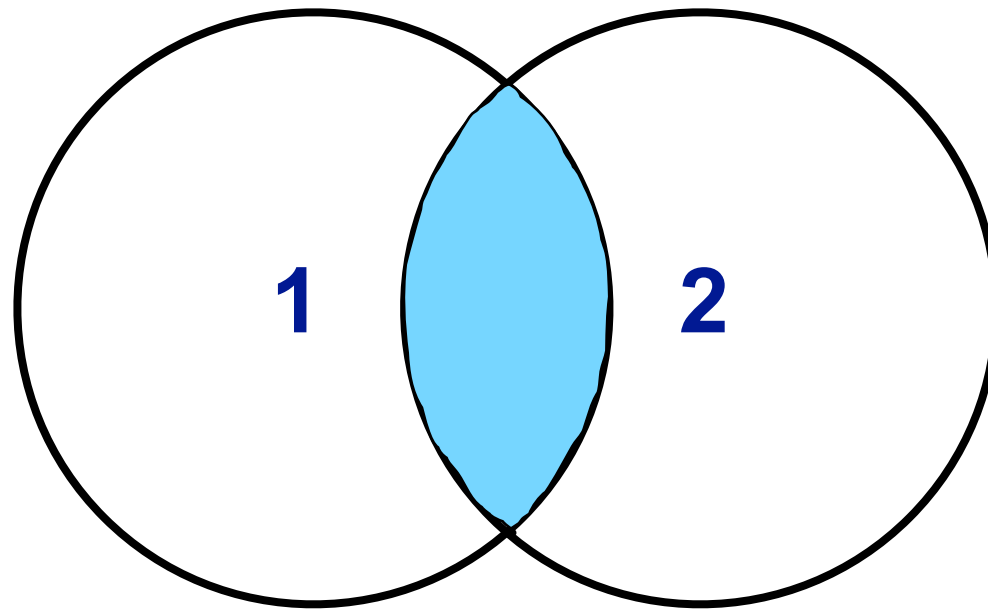
Search

[See more...](#)

Recent activity

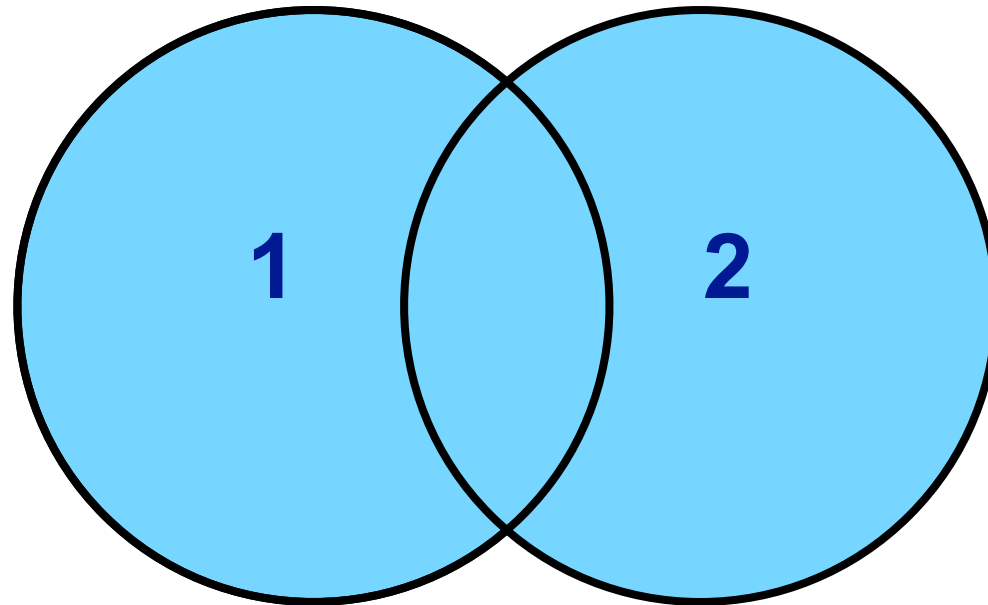
[Turn Off](#) [Clear](#)

1 AND 2



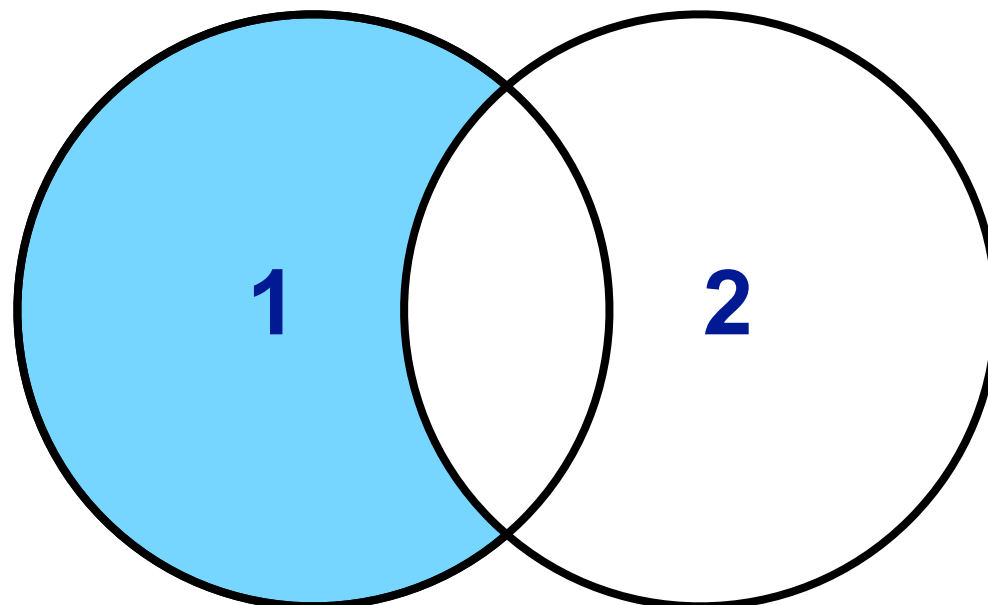
**ras AND disease
(1185 results)**

1 OR 2



**ras OR disease
(134,872 results)**

1 NOT 2



**ras NOT disease
(84,448 results)**

NCBI Resources How To Sign in to NCBI

Gene Gene (ras) AND "Homo sapiens"[porgn:__txid9606] Search

Save search Advanced Help

Show additional filters

Display Settings: Tabular, 20 per page, Sorted by Relevance Send to: Hide sidebar >>

Results: 1 to 20 of 1126 << First < Prev Page 1 of 57 Next > Last >>

Filters activated: Current only. Clear all to show 1499 items.

Name/Gene ID	Description	Location	Aliases
<input type="checkbox"/> NRAS ID: 4893	neuroblastoma RAS viral (v-ras) oncogene homolog [Homo sapiens (human)]	Chromosome 1, NC_000001.11 (114704464..114716894, complement)	RP5-1000E10.2, ALPS4, CMNS, N-ras, NCMS1, NS6, NRAS
<input type="checkbox"/> KRAS ID: 3845	Kirsten rat sarcoma viral oncogene homolog [Homo sapiens (human)]	Chromosome 12, NC_000012.12 (25205246..25250923, complement)	C-K-RAS, CFC2, K-RAS2A, K-RAS2B, K-RAS4A, K-RAS4B, KI-RAS1, KRAS2, NS, NS3, RASK2

Gene sources

Genomic

Categories

Alternatively spliced

Annotated genes

Non-coding

Protein-coding

Pseudogene

Sequence content

CCDS

Ensembl

RefSeq

Status clear

✓ Current only

Chromosome locations

Select

Filters: Manage Filters

Find related data

Database: Select

Find items

Search details

ras[All Fields] AND "Homo sapiens"[porgn] AND alive[property]

Search See more...

Recent activity

Turn Off Clear

KRAS Kirsten rat sarcoma

www.ncbi.nlm.nih.gov/gene/3845

NCBI

Resources

How To

Sign in to NCBI

Gene

Gene

Search

Advanced

Help

Display Settings: Full Report

Send to:

Hide sidebar >>

KRAS Kirsten rat sarcoma viral oncogene homolog [*Homo sapiens* (human)]

Gene ID: 3845, updated on 4-Jan-2015

Summary

Official Symbol

Official Full Name

Primary source

See related

Gene type

RefSeq status

Organism

Lineage

Also known as

KRAS provided by HGNC

Kirsten rat sarcoma viral oncogene homolog provided by HGNC

HGNC:HGNC:6407

Ensembl:ENSG00000133703; HPRD:01817; MIM:190070; Vega:OTTHUMG00000171193

protein coding

REVIEWED

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

NS; NS3; CFC2; KRAS1; KRAS2; RASK2; KI-RAS; C-K-RAS; K-RAS2A; K-

Table of contents

Summary

Genomic context

Genomic regions, transcripts, and products

Bibliography

Phenotypes

Variation

HIV-1 interactions

Pathways from BioSystems

Interactions

General gene information

Markers, Related pseudogene(s), Homology, Gene Ontology

General protein information

NCBI Reference Sequences (RefSeq)

Example Questions:

What chromosome location and what genes are in the vicinity?

Gene

[Display Summary](#)

KRAS
(human)

Gene ID: 3845, updated on 4-Jan-2015

Summary

Official Symbol KRAS provided by [HGNC](#)
Official Full Name Kirsten rat sarcoma viral oncogene homolog provided by [HGNC](#)
Primary source [HGNC:HGNC:6407](#)
See related [Ensembl:ENSG00000133703](#); [HPRD:01817](#); [MIM:190070](#);
[Vega:OTTHUMG00000171193](#)
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 NS; NS3; CFC2; KRAS1; KRAS2; RASK2; KI-RAS; C-K-RAS; K-RAS2A; K-

[Sign in to NCBI](#)

[Search](#)

[Help](#)

[Hide sidebar >>](#)

Table of contents

[Summary](#)

[Genomic context](#)

[Genomic regions, transcripts, and products](#)

[Bibliography](#)

[Phenotypes](#)

[Variation](#)

[HIV-1 interactions](#)

[Pathways from BioSystems](#)

[Interactions](#)

[General gene information](#)

[Markers, Related pseudogene\(s\),
Homology, Gene Ontology](#)

[General protein information](#)

[NCBI Reference Sequences
\(RefSeq\)](#)

KRAS Kirsten rat sarcoma

www.ncbi.nlm.nih.gov/gene/3845#genomic-context

Genomic context

Location: 12p12.1

Exon count: 6

See KRAS in [Epigenomics](#), [MapView](#)

Annotation release	Status	Assembly	Chr	Location
106	current	GRCh38 (GCF_000001405.26)	12	NC_000012.12 (25205246..25250923, complement)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	12	NC_000012.11 (25358180..25403870, complement)

Chromosome 12 - NC_000012.12

Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence: NC_000012.12 chromosome 12 reference GRCh38 Primary Assembly

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

- BioAssay by Target (List)
- BioAssay by Target (Summary)
- BioAssay, by Gene target
- BioAssays, RNAi Target, Active
- BioAssays, RNAi Target, Tested
- BioProjects
- BioSystems
- Books
- CCDS
- ClinVar
- Conserved Domains
- dbVar
- EST
- Full text in PMC
- Full text in PMC_nucleotide
- Gene neighbors
- Genome
- GEO Profiles
- GTR
- HomoloGene
- Map Viewer
- MedGen
- Nucleotide

Example Questions:
What 'molecular functions', 'biological processes', and 'cellular component' information is available?

Official Symbol KRAS provided by [HGNC](#)
Official Full Name Kirsten rat sarcoma viral oncogene homolog provided by [HGNC](#)
Primary source [HGNC:HGNC:6407](#)
See related [Ensembl:ENSG00000133703](#); [HPRD:01817](#); [MIM:190070](#); [Vega:OTTHUMG00000171193](#)
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 NS; NS3; CFC2; KRAS1; KRAS2; RASK2; KI-RAS; C-K-RAS; K-RAS2A; K-

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Phenotypes
- Variation
- HIV-1 interactions
- Pathways from BioSystems
- Interactions
- General gene information**
 - Markers, Related pseudogene(s), Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences

Gene Ontology [Provided by GOA](#)

Function	Evidence Code	Pubs
GDP binding	IEA	
GMP binding	IEA	
GTP binding	IEA	
LRR domain binding	IEA	
protein binding	IPI	PubMed
protein complex binding	IDA	PubMed

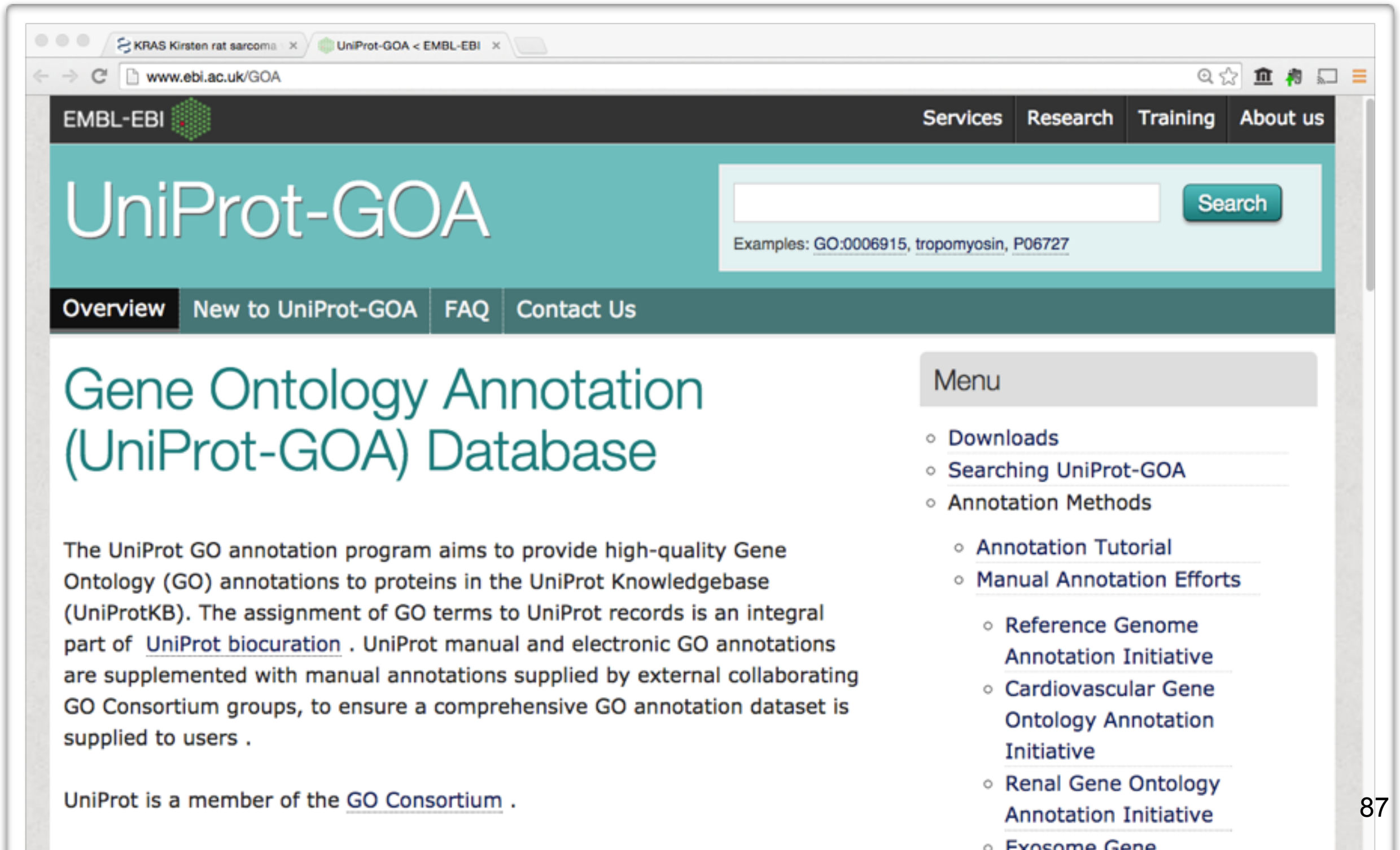
Items 1 - 25 of 33 < Prev Page 1 of 2 Next >

Process	Evidence Code	Pubs
Fc-epsilon receptor signaling pathway	TAS	
GTP catabolic process	IEA	
MAPK cascade	TAS	
Ras protein signal transduction	TAS	
actin cytoskeleton organization	IEA	
activation of MAPKK activity	TAS	
axon guidance	TAS	
blood coagulation	TAS	



GO: Gene Ontology

GO provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data



The screenshot shows the UniProt-GOA website interface. At the top, there's a browser window with two tabs: 'KRAS Kirsten rat sarcoma' and 'UniProt-GOA < EMBL-EBI'. The address bar shows 'www.ebi.ac.uk/GOA'. The website header includes the EMBL-EBI logo and navigation links for 'Services', 'Research', 'Training', and 'About us'. Below the header, the 'UniProt-GOA' logo is prominently displayed on the left, and a search bar with a 'Search' button is on the right. Below the search bar, there are example search terms: 'Examples: GO:0006915, tropomyosin, P06727'. A secondary navigation bar contains links for 'Overview', 'New to UniProt-GOA', 'FAQ', and 'Contact Us'. The main content area features the title 'Gene Ontology Annotation (UniProt-GOA) Database' in large teal text. Below this, a paragraph describes the UniProt GO annotation program's goals and methods. To the right of the main text, a 'Menu' section lists various resources: 'Downloads', 'Searching UniProt-GOA', 'Annotation Methods', 'Annotation Tutorial', 'Manual Annotation Efforts', 'Reference Genome Annotation Initiative', 'Cardiovascular Gene Ontology Annotation Initiative', 'Renal Gene Ontology Annotation Initiative', and 'Exosome Gene'.

EMBL-EBI

Services Research Training About us

UniProt-GOA

Search

Examples: [GO:0006915](#), [tropomyosin](#), [P06727](#)

Overview New to UniProt-GOA FAQ Contact Us

Gene Ontology Annotation (UniProt-GOA) Database

The UniProt GO annotation program aims to provide high-quality Gene Ontology (GO) annotations to proteins in the UniProt Knowledgebase (UniProtKB). The assignment of GO terms to UniProt records is an integral part of [UniProt biocuration](#). UniProt manual and electronic GO annotations are supplemented with manual annotations supplied by external collaborating GO Consortium groups, to ensure a comprehensive GO annotation dataset is supplied to users.

UniProt is a member of the [GO Consortium](#).

Menu

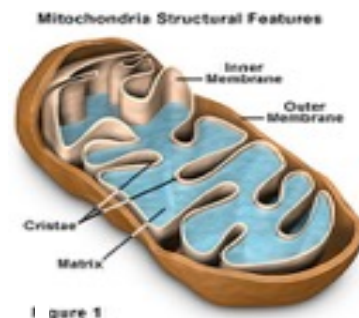
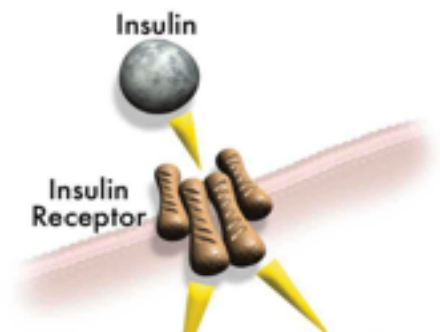
- Downloads
- Searching UniProt-GOA
- Annotation Methods
- Annotation Tutorial
- Manual Annotation Efforts
 - Reference Genome Annotation Initiative
 - Cardiovascular Gene Ontology Annotation Initiative
 - Renal Gene Ontology Annotation Initiative
 - Exosome Gene

Why do we need Ontologies?

- Annotation is essential for capturing the understanding and knowledge associated with a sequence or other molecular entity
- Annotation is traditionally recorded as “free text”, which is easy to read by humans, but has a number of disadvantages, including:
 - ▶ Difficult for computers to parse
 - ▶ Quality varies from database to database
 - ▶ Terminology used varies from annotator to annotator
- Ontologies are annotations using standard vocabularies that try to address these issues
- GO is integrated with UniProt and many other databases including a number at NCBI

GO Ontologies

- There are three ontologies in GO:
 - ▶ **Biological Process**
A commonly recognized series of events
e.g. cell division, mitosis,
 - ▶ **Molecular Function**
An elemental activity, task or job
e.g. kinase activity, insulin binding
 - ▶ **Cellular Component**
Where a gene product is located
e.g. mitochondrion, mitochondrial
membrane



Gene Ontology

Provided by GOA

Function	Evidence Code	Pubs
GDP binding		
GMP binding		
GTP binding		
LRR domain binding		
protein binding		
protein complex binding		

Process	Code	Pubs
Fc-epsilon receptor signaling pathway	TAS	
GTP catabolic process	IEA	
MAPK cascade	TAS	
Ras protein signal transduction	TAS	
actin cytoskeleton organization	IEA	
activation of MAPKK activity	TAS	
axon guidance	TAS	
blood coagulation	TAS	

The 'Gene Ontology' or GO is actually maintained by the EBI so lets switch or link over to UniProt also from the EBI.

⋮

↓

Scroll down to UniProt link

UniProt will detail much more information for protein coding genes such as this one

The screenshot shows the NCBI Gene page for KRAS (3845). The browser address bar displays www.ncbi.nlm.nih.gov/gene/3845#gene-ontology. The page header includes genomic coordinates [X01669.1](#) and [CAA25828.1](#). A pagination bar indicates 'Items 1 - 25 of 43' on 'Page 1 of 2'. A table lists protein accessions and their corresponding links:

Protein Accession	Links	
	GenPept Link	UniProtKB Link
P01116.1	GenPept	UniProtKB/Swiss-Prot:P01116

The UniProtKB link is highlighted with a red box. A red arrow points to it with the text 'Scroll down to UniProt link'. Below the table is an 'Additional links' section. The footer contains navigation links organized into five columns: GETTING STARTED, RESOURCES, POPULAR, FEATURED, and NCBI INFORMATION.

You are here: NCBI > Genes & Expression > Gene [Write to the Help Desk](#)

GETTING STARTED	RESOURCES	POPULAR	FEATURED	NCBI INFORMATION
NCBI Education	Chemicals & Bioassays	PubMed	Genetic Testing Registry	About NCBI
NCBI Help Manual	Data & Software	Bookshelf	PubMed Health	Research at NCBI
NCBI Handbook	DNA & RNA	PubMed Central	GenBank	NCBI News
Training & Tutorials	Domains & Structures	PubMed Health	Reference Sequences	NCBI FTP Site
	Genes & Expression	BLAST	Gene Expression Omnibus	NCBI on Facebook
	Genetics & Medicine	Nucleotide	Map Viewer	NCBI on Twitter
	Genomes & Maps	Genome	Human Genome	NCBI on YouTube
	Homology	SNP	Mouse Genome	
	Literature	Gene	Influenza Virus	
	Proteins	Protein	Primer-BLAST	
	Sequence Analysis	PubChem	Sequence Read Archive	
	Taxonomy			

UniProt will detail much more information for protein coding genes

UniProtKB

Advanced

BLAST Align Retrieve/ID Mapping Help Contact

Basket

P01116 - RASK_HUMAN

Protein | **GTPase KRas**

Gene | **KRAS**

Organism | *Homo sapiens (Human)*

Status | **Reviewed** - ●●●●● - Experimental evidence at protein levelⁱ

BLAST Align Format Add to basket History Feedback Help video

Functionⁱ

Ras proteins bind GDP/GTP and possess intrinsic GTPase activity. Plays an important role in the regulation of cell proliferation (PubMed:23698361, PubMed:22711838). 2 Publications Curated

Enzyme regulationⁱ

Alternates between an inactive form bound to GDP and an active form bound to GTP. Activated by a guanine nucleotide-exchange factor (GEF) and inactivated by a GTPase-activating protein (GAP). Interaction with SOS1 promotes exchange of bound GDP by GTP. 3 Publications

Regions

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Nucleotide binding ⁱ	10 – 18	9	GTP 2 Publications			
Nucleotide binding ⁱ	29 – 35	7	GTP 2 Publications			
Nucleotide binding ⁱ	59 – 60	2	GTP 2 Publications			

Display None

- ☒ FUNCTION
- ☒ NAMES & TAXONOMY
- ☒ SUBCELL. LOCATION
- ☒ PATHOL./BIOTECH
- ☒ PTM / PROCESSING
- ☒ EXPRESSION
- ☒ INTERACTION
- ☒ STRUCTURE
- ☒ FAMILY & DOMAINS
- ☒ SEQUENCES (2)
- ☒ CROSS-REFERENCES

UniProt

UniProtKB

Advanced

BLASTAlignRetrieve/ID MappingHelpContact

Basket

P01116 - RASK_HUMAN

Protein

Gene

Organism

Status

GTPase KRas

KRAS

Homo sapiens (Human)

Reviewed -

Display

None

FUNCTION

NAMES & TAXONOMY

SUBCELL. LOCATION

PATHOL./BIOTECH

PTM / PROCESSING

EXPRESSION

INTERACTION

STRUCTURE

FAMILY & DOMAINS

SEQUENCES (2)

CROSS-REFERENCES

BLAST

Align

Download

Function

Ras proteins bind GDP/GTP and possess intrinsic GTPase activity. Plays an important role in the regulation of cell proliferation (PubMed:23698361, PubMed:22711838). 2 Publications Curated

Enzyme regulation

Alternates between an inactive form bound to GDP and an active form bound to GTP. Activated by a guanine nucleotide-exchange factor (GEF) and inactivated by a GTPase-activating protein (GAP). Interaction with SOS1 promotes exchange of bound GDP by GTP. 3 Publications

Regions

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Nucleotide binding	10 – 18	9	GTP 2 Publications			
Nucleotide binding	29 – 35	7	GTP 2 Publications			
Nucleotide binding	59 – 60	2	GTP 2 Publications			

Example Questions:

What positions in the protein are responsible for GTP binding?

Example Questions:

What variants of this enzyme are involved in gastric cancer and other human diseases?

Display **None**

- ☒ FUNCTION
- ☒ NAMES & TAXONOMY
- ☒ SUBCELL. LOCATION
- ☒ **PATHOL./BIOTECH**
- ☒ PTM / PROCESSING
- ☒ EXPRESSION
- ☒ INTERACTION
- ☒ STRUCTURE
- ☒ FAMILY & DOMAINS
- ☒ SEQUENCES (2)
- ☒ CROSS-REFERENCES
- ☒ PUBLICATIONS
- ☒ ENTRY INFORMATION
- ☒ MISCELLANEOUS
- ☒ SIMILAR PROTEINS

[Top](#)

Pathology & Biotechⁱ

Involvement in diseaseⁱ

LEUKEMIA, ACUTE MYELOGENOUS (AML)

[MIM:601626]: A subtype of acute leukemia, a cancer of the white blood cells. AML is a malignant disease of bone marrow characterized by maturational arrest of hematopoietic precursors at an early stage of development. Clonal expansion of myeloid blasts occurs in bone marrow, blood, and other tissue. Myelogenous leukemias develop from changes in cells that normally produce neutrophils, basophils, eosinophils and monocytes. [1 Publication](#)

Note: The disease is caused by mutations affecting the gene represented in this entry.

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Natural variant ⁱ	10 - 10	1	G → GG in one individual with AML; expression in 3T3 cell causes cellular transformation; expression in COS cells activates the Ras-MAPK signaling pathway; lower GTPase activity; faster GDP dissociation rate. 1 Publication		VAR_034601	

LEUKEMIA, JUVENILE MYELOMONOCYTIC (JMML)

[MIM:607785]: An aggressive pediatric myelodysplastic syndrome/myeloproliferative disorder characterized by malignant transformation in the hematopoietic stem cell compartment with proliferation of differentiated progeny. Patients have splenomegaly, enlarged lymph nodes, rashes, and hemorrhages. Note: The disease is caused by mutations affecting the gene represented in this entry.

NOONAN SYNDROME 3 (NS3)

[MIM:609942]: A form of Noonan syndrome, a disease characterized by short stature, facial dysmorphic features such as hypertelorism, a downward eyeslant and low-set posteriorly rotated ears, and a high incidence of congenital heart

Example Questions:

Are high resolution protein structures available to examine the details of these mutations?

Display None Structure¹

☒ FUNCTION
☒ NAMES & TAXONOMY
☒ SUBCELL. LOCATION
☒ PATHOL./BIOTECH
☒ PTM / PROCESSING
☒ EXPRESSION
☒ INTERACTION
☒ **STRUCTURE**
☒ FAMILY & DOMAINS
☒ SEQUENCES (2)
☒ CROSS-REFERENCES
☒ PUBLICATIONS
☒ ENTRY INFORMATION
☒ MISCELLANEOUS
☒ SIMILAR PROTEINS

▲ Top


Secondary structure
1
Legend: Helix Turn Beta strand 189

Show more details

3D structure databases

Select the link destinations:
☒ PDBeⁱ
☐ RCSB PDBⁱ
☐ PDBjⁱ

Entry	Method	Resolution (Å)	Chain	Positions	PDBsum
1D8D	X-ray	2.00	P	178-188	[*]
1D8E	X-ray	3.00	P	178-188	[*]
1KZO	X-ray	2.20	C	169-173	[*]
1KZP	X-ray	2.10	C	169-173	[*]
3GFT	X-ray	2.27	A/B/C/D/E/F	1-164	[*]
4DSN	X-ray	2.03	A	2-164	[*]
4DSO	X-ray	1.85	A	2-164	[*]
4EPR	X-ray	2.00	A	1-164	[*]
4EPT	X-ray	2.00	A	1-164	[*]
4EPV	X-ray	1.35	A	1-164	[*]
4EPW	X-ray	1.70	A	1-164	[*]
4EPX	X-ray	1.76	A	1-164	[*]
4EPY	X-ray	1.80	A	1-164	[*]
4L8G	X-ray	1.52	A	1-169	[*]
4LDJ	X-ray	1.15	A	1-164	[*]
4LPK	X-ray	1.50	A/B	1-169	[*]



Example Questions:

What is known about the protein family, its species distribution, number in humans and residue-wise conservation, etc... ?

The screenshot shows the UniProt website for the KRAS protein (P01116). The left sidebar contains a list of categories with checkboxes, and the 'Family & Domains' category is highlighted with a red box. The main content area displays a table of family and domain databases, with the 'Pfam' entry (PF00071) highlighted by an orange box. A text box on the right states 'PFAM is one of the best protein family databases'.

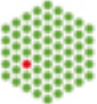
Database	Hit
Gene3D ⁱ	3.40.50.300. 1 hit.
InterPro ⁱ	IPR027417. P-loop_NTPase. IPR005225. Small_GTP-bd_dom. IPR001806. Small_GTPase. IPR020849. Small_GTPase_Ras. [Graphical view]
PANTHER ⁱ	PTHR24070. PTHR24070. 1 hit
Pfam ⁱ	PF00071. Ras. 1 hit. [Graphical view]
PRINTS ⁱ	PR00449. RASTRNSFRMNG.
SMART ⁱ	SM00173. RAS. 1 hit. [Graphical view]
SUPFAM ⁱ	SSF52540. SSF52540. 1 hit.
TIGRFAMs ⁱ	TIGR00231. small_GTP. 1 hit.
PROSITE ⁱ	PS51421. RAS. 1 hit. [Graphical view]

Sequences (2)ⁱ

Sequence statusⁱ: Complete.
Sequence processingⁱ: The displayed sequence is further processed into a mature form.
This entry describes 2 isoformsⁱ produced by **alternative splicing**. [Align](#)

Example Questions:

What is known about the protein family, its **species distribution**, number in humans and residue-wise conservation, etc... ?

EMBL-EBI  HOME | SEARCH

Family: Ras (PF00071)

332 architectures 21243 sequences 30 interactions 1006 species 663 structures

- Summary
- Domain organisation
- Clan
- Alignments
- HMM logo
- Trees
- Curation & model
- Species**
- Interactions
- Structures

Jump to...

Summary: Ras family

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

[Wikipedia: Ras subfamily](#) [Wikipedia: Ras superfamily](#) [Pfam](#) [InterPro](#)

This is the Wikipedia entry entitled "[Ras subfamily](#)". [More...](#)

Ras subfamily [Edit Wikipedia article](#)

This article is about p21/Ras protein. For the p21/waf1 protein, see [p21](#).

Ras is the name given to a [family of related proteins](#) which is ubiquitously expressed in all cell lineages and organs. All Ras protein family members belong to a class of protein called [small GTPase](#), and are involved in transmitting signals within cells ([cellular signal transduction](#)). Ras is the prototypical member of the [Ras superfamily](#) of proteins, which are all related in 3D structure and regulate diverse cell behaviours.

The name 'Ras' is an abbreviation of 'Rat [sarcoma](#)', reflecting the way the first members of the protein family were discovered. The name ras is also used to refer to the family of [genes](#) encoding those proteins.

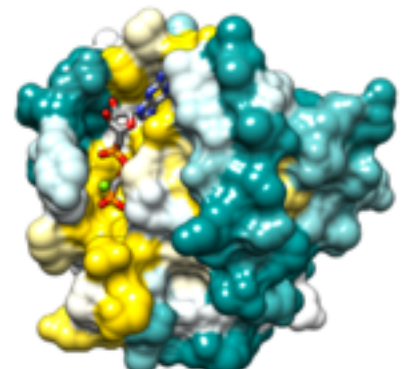
When Ras is 'switched on' by incoming signals, it subsequently switches on other proteins, which ultimately turn on genes involved in [cell growth](#), [differentiation](#) and [survival](#). As a result, mutations in ras genes can lead to the production of permanently activated Ras proteins. This can cause unintended and overactive signalling inside the cell, even in the absence of incoming signals.

Because these signals result in cell growth and division, overactive Ras signaling can ultimately lead to [cancer](#).^[1] The 3 Ras genes in humans ([HRAS](#), [KRAS](#), and [NRAS](#)) are the most common [oncogenes](#) in human [cancer](#); mutations that permanently activate Ras are found in 20% to 25% of all human tumors and up to 90% in certain types of cancer (e.g., [pancreatic cancer](#)).

^[2] For this reason, Ras inhibitors are being studied as a treatment for cancer, and other diseases with Ras overexpression.

Contents [\[hide\]](#)

- History
- Structure
- Function
 - 3.1 Activation and deactivation
 - 3.2 Membrane attachment
- Members
- Ras in cancer
 - 5.1 Inappropriate activation
 - 5.2 Constitutively active Ras

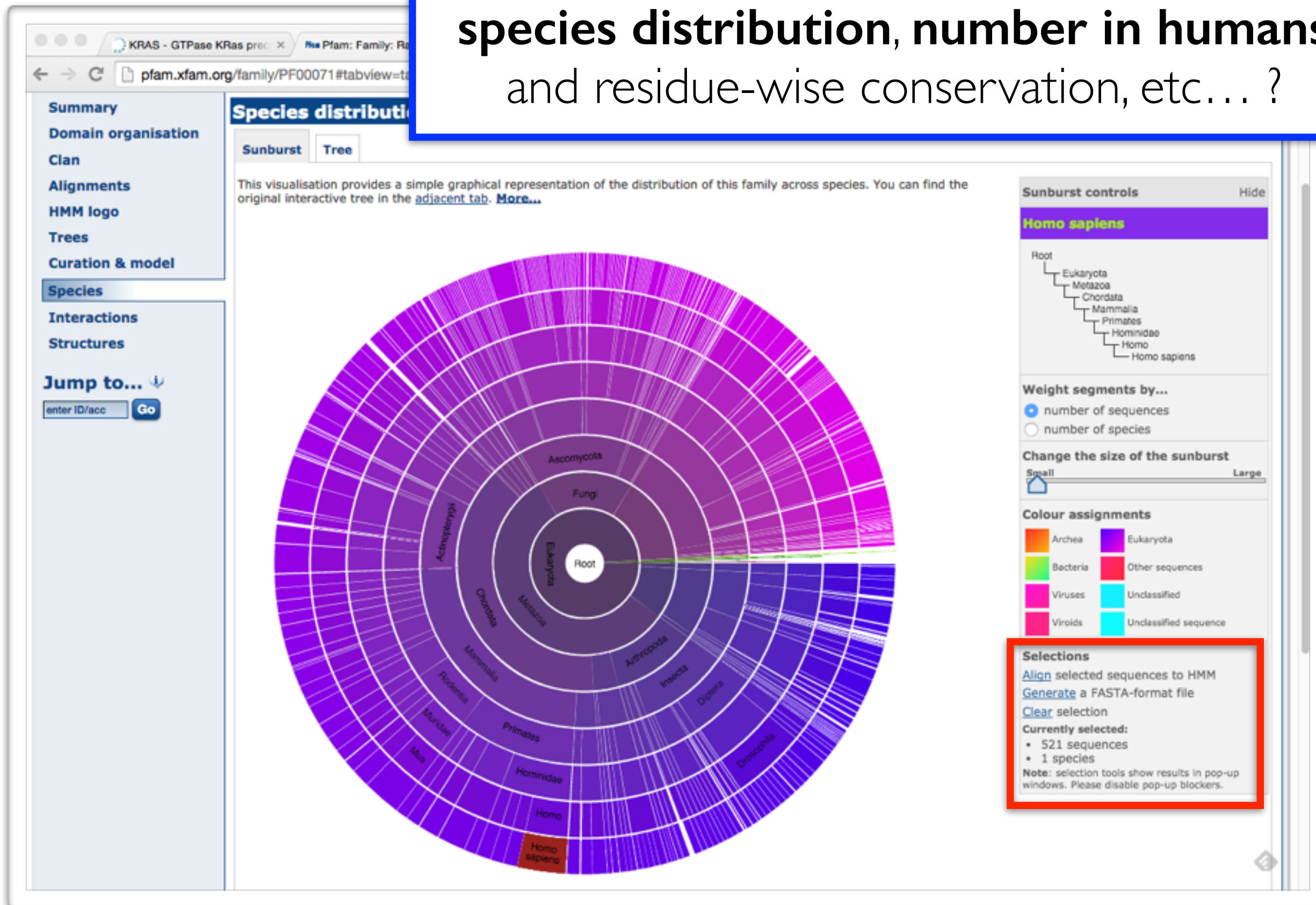


H-Ras structure PDB 121p, surface colored by conservation in Pfam seed alignment: gold, most conserved; dark cyan, least conserved.

Identifiers	
Symbol	Ras
Pfam	PF00071
InterPro	IPR013753
PROSITE	PDOC00017
SCOP	Sp21
SUPERFAMILY	Sp21

Example Questions:

What is known about the protein family, its **species distribution, number in humans** and residue-wise conservation, etc... ?



Example Questions:

What is known about the protein family, its species distribution, number in humans and **residue-wise conservation**, etc... ?

The screenshot displays the Pfam website interface for the PF00071 protein family. The main navigation bar includes links for Summary, Domain organisation, Clan, Alignment, HMM log, Trees, Curation, Species, Interact, and Structure. The Species distribution section shows a sunburst chart for Homo sapiens, with a tree view showing the hierarchy from Root to Eukaryota, Metazoa, Chordata, Mammalia, Primates, Hominidae, and Homo sapiens. The Alignment viewer shows a multiple sequence alignment of 536 sequences, with a selection tool on the right. The selection tool includes options for Align, Generate, and Clear, and a list of currently selected sequences (521 sequences, 1 species). The alignment viewer also shows a table of sequence identifiers and their corresponding amino acid sequences, with a color-coded background for each sequence.

Species distribution

Homo sapiens

Root
Eukaryota
Metazoa
Chordata
Mammalia
Primates
Hominidae
Homo
Homo sapiens

Weight segments by...

☒ number of sequences
☐ number of species

Change the size of the sunburst

Small Large

Colour assignments

Archea Eukaryota
Bacteria Other sequences
Viruses Unclassified
Viroids Unclassified sequence

Selections

[Align](#) selected sequences to HMM
[Generate](#) a FASTA-format file
[Clear](#) selection

Currently selected:

- 521 sequences
- 1 species

Note: selection tools show results in pop-up windows. Please disable pop-up blockers.

Alignment for selected sequences

Currently showing rows 1 to 30 of 536 rows in this alignment. Show 30 rows of alignment

EMBL-EBI

PF00071

Alignment viewer

PF00071/alignment/view?jobId=EDCA403E-9836-11E4-B360-10B3298E2F76

1 2 3 4 5 6 7 8 9 10 11 ...

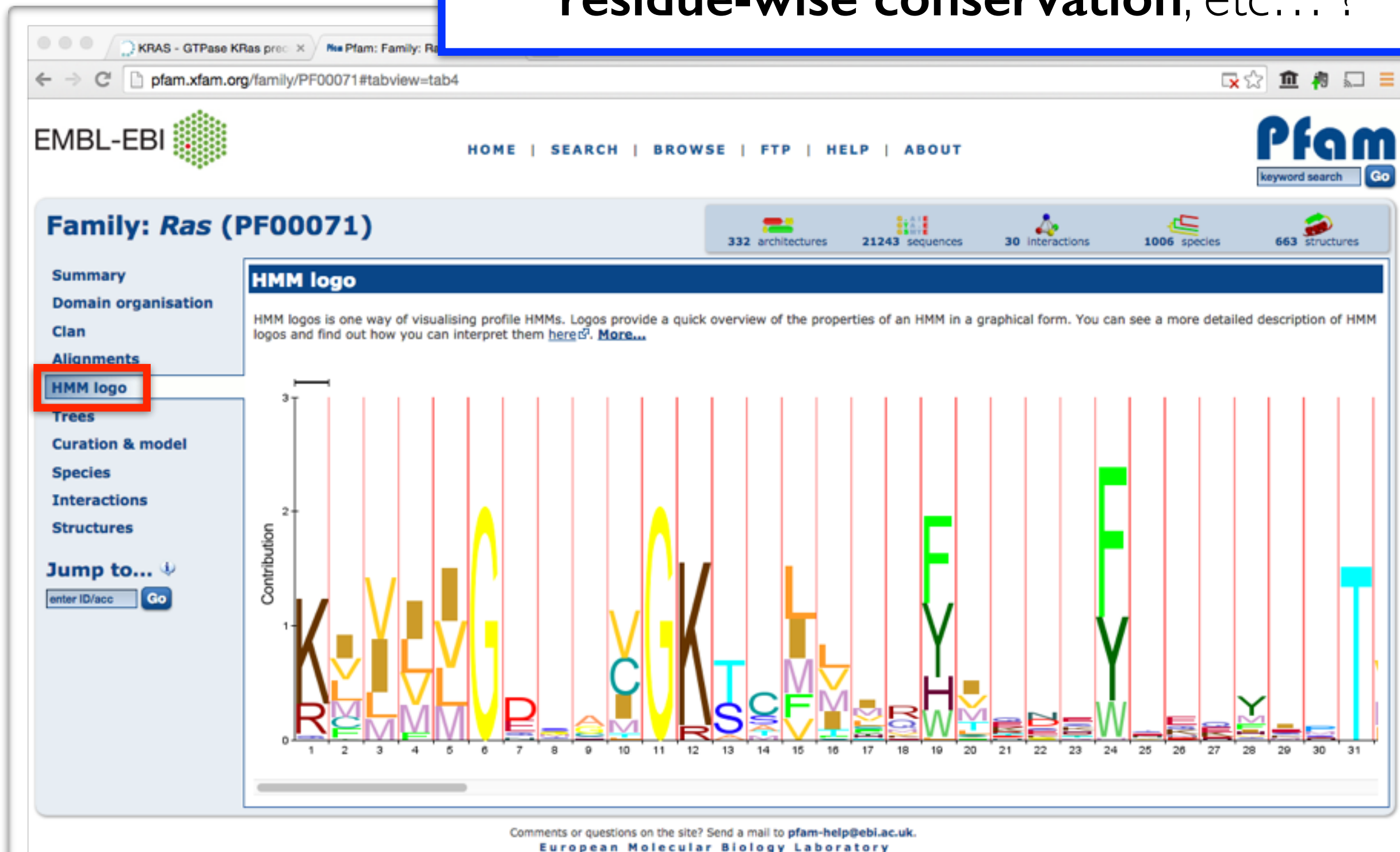
There are 18 pages in this alignment. Show page 1

[Download this alignment.](#)

[Close window](#)

Example Questions:

What is known about the protein family, its species distribution, number in humans and **residue-wise conservation**, etc... ?



Family: *Kinesin* (PF00225)

⌂ Loading page components (1 remaining)...

 126 architectures  4150 sequences  6 interactions  248 species  114 structures

[Summary](#)[Domain organisation](#)[Clans](#)[Alignments](#)[HMM logo](#)[Trees](#)[Curation & models](#)[Species](#)[Interactions](#)[Structures](#)

Jump to... ⓘ

[Go](#)

Interactions

There are **6** interactions for this family. [More...](#)

[Tubulin](#)
[Tubulin_C](#)

[Tubulin_C](#)

[Kinesin](#)

[Tubulin](#)

[Kinesin](#)

Questions or comments: pfam@janelia.hhmi.org

Howard Hughes Medical Institute

Family: *Kinesin* (PF00225)

 126 architectures
  4150 sequences
  6 interactions
  248 species
  114 structures

Summary

Domain organisation

Clans

Alignments

HMM logo

Trees

Curation & models

Species

Interactions

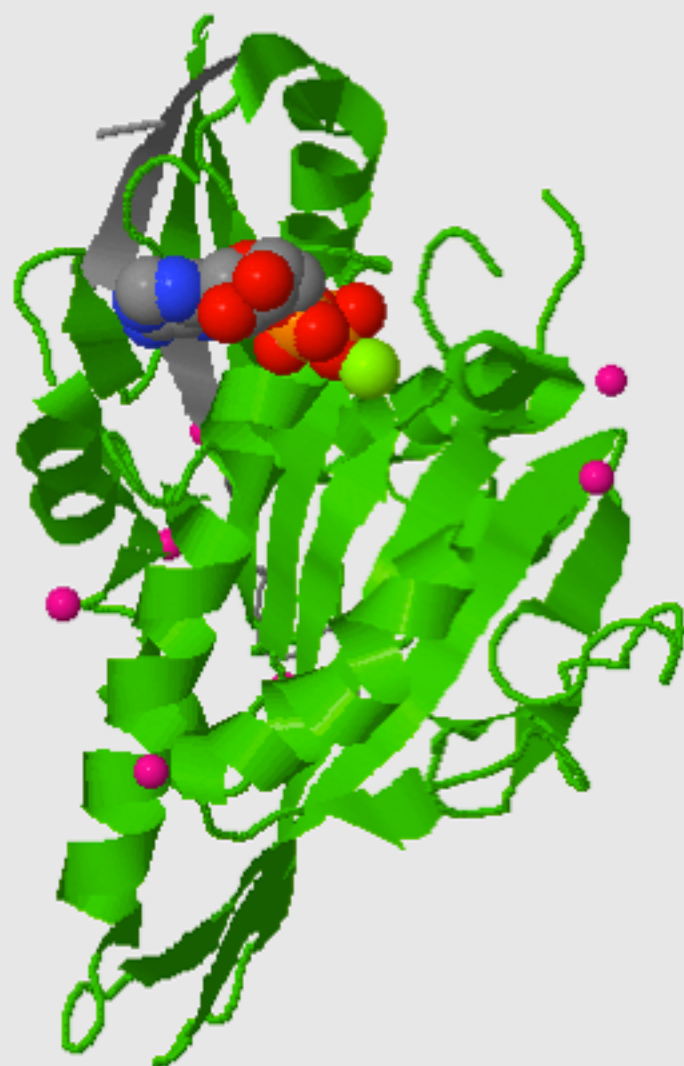
Structures

Jump to...

Structures

For those sequences which have a structure in the [Protein DataBank](#), we use the mapping between [UniProt](#), PDB and Pfam coordinate systems from the [PDB](#) group, to allow us to map Pfam domains onto UniProt sequences and three-dimensional protein structures. The table below shows the structures on which the **Kinesin** domain has been found.

UniProt entry	UniProt residues	PDB ID	PDB chain ID	PDB residues	View
A8BKD1_GIALA	11 - 335	2vvq	A	11 - 335	Jmol AstexViewer SPICE
			B	11 - 335	Jmol AstexViewer SPICE
CENPE_HUMAN	12 - 329	1t5c	A	12 - 329	Jmol AstexViewer SPICE
			B	12 - 329	Jmol AstexViewer SPICE
KAR3_YEAST	392 - 723	1f9t	A	392 - 723	Jmol AstexViewer SPICE
			A	392 - 723	Jmol AstexViewer SPICE
			A	392 - 723	Jmol AstexViewer SPICE
			A	392 - 723	Jmol AstexViewer SPICE
			B	392 - 723	Jmol AstexViewer SPICE
			A	392 - 723	Jmol AstexViewer SPICE
KI13B_HUMAN	11 - 352	3qbj	A	11 - 352	Jmol AstexViewer SPICE
			B	11 - 352	Jmol AstexViewer SPICE
			C	11 - 352	Jmol AstexViewer SPICE
		1ii6	A	24 - 359	Jmol AstexViewer SPICE
			B	24 - 359	Jmol AstexViewer SPICE
		1q0b	A	24 - 359	Jmol AstexViewer SPICE
			B	24 - 359	Jmol AstexViewer SPICE
		1x88	A	24 - 359	Jmol AstexViewer SPICE
			B	24 - 359	Jmol AstexViewer SPICE
			A	24 - 359	Jmol AstexViewer SPICE
			A	24 - 359	Jmol AstexViewer SPICE



Jmol

PDB			UniProt			Pfam family	Colour
Chain	Start	End	ID	Start	End		
A	49	368	KIF22_HUMAN	49	368	Kinesin (PF00225)	

ENTREZ & BLAST:

TOOLS FOR SEARCHING AND ACCESSING
MOLECULAR DATA AT NCBI

Entrez: Integrated search of NCBI databases

The image shows the NCBI Entrez homepage. On the left, there is a vertical navigation menu with categories like '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', and 'Variation'. A dropdown menu is open from the 'Resource List (A-Z)' category, displaying a list of databases: 'All Databases' (checked), 'PubMed', 'Protein', 'Nucleotide', 'GSS', 'EST', 'Structure', 'Genome', 'BioProject', 'BioSample', 'BioSystems', 'Books', 'Conserved Domains', 'Clone', 'dbGaP', 'dbVar', 'Epigenomics', 'Gene', 'GEO DataSets', 'GEO Profiles', 'HomoloGene', 'MeSH', 'NCBI Web Site', 'NLM Catalog', 'OMIA', 'OMIM', 'PMC', 'PopSet', 'Probe', 'Protein Clusters', 'PubChem BioAssay', 'PubChem Compound', 'PubChem Substance', 'PubMed Health', 'SNP', 'SRA', 'Taxonomy', 'ToolKit', 'ToolKitAll', 'UniGene', and 'UniSTS'. The main content area features a 'Welcome to NCBI' message, a search bar with a 'Search' button, and a 'Popular Resources' section listing 'PubMed', 'Bookshelf', 'PubMed Central', 'PubMed Health', 'BLAST', 'Nucleotide', 'Genome', 'SNP', 'Gene', and 'Protein'. Below this is an 'NCBI Announcements' section with news items dated May 2012. A text box in the center of the page states: 'Entrez is available from the main NCBI homepage or from the homepage of individual databases'.

NCBI
National Center for
Biotechnology Information

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

✓ All Databases
PubMed
Protein
Nucleotide
GSS
EST
Structure
Genome
BioProject
BioSample
BioSystems
Books
Conserved Domains
Clone
dbGaP
dbVar
Epigenomics
Gene
GEO DataSets
GEO Profiles
HomoloGene
MeSH
NCBI Web Site
NLM Catalog
OMIA
OMIM
PMC
PopSet
Probe
Protein Clusters
PubChem BioAssay
PubChem Compound
PubChem Substance
PubMed Health
SNP
SRA
Taxonomy
ToolKit
ToolKitAll
UniGene
UniSTS

Welcome to NCBI
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
[About NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS Feeds](#)

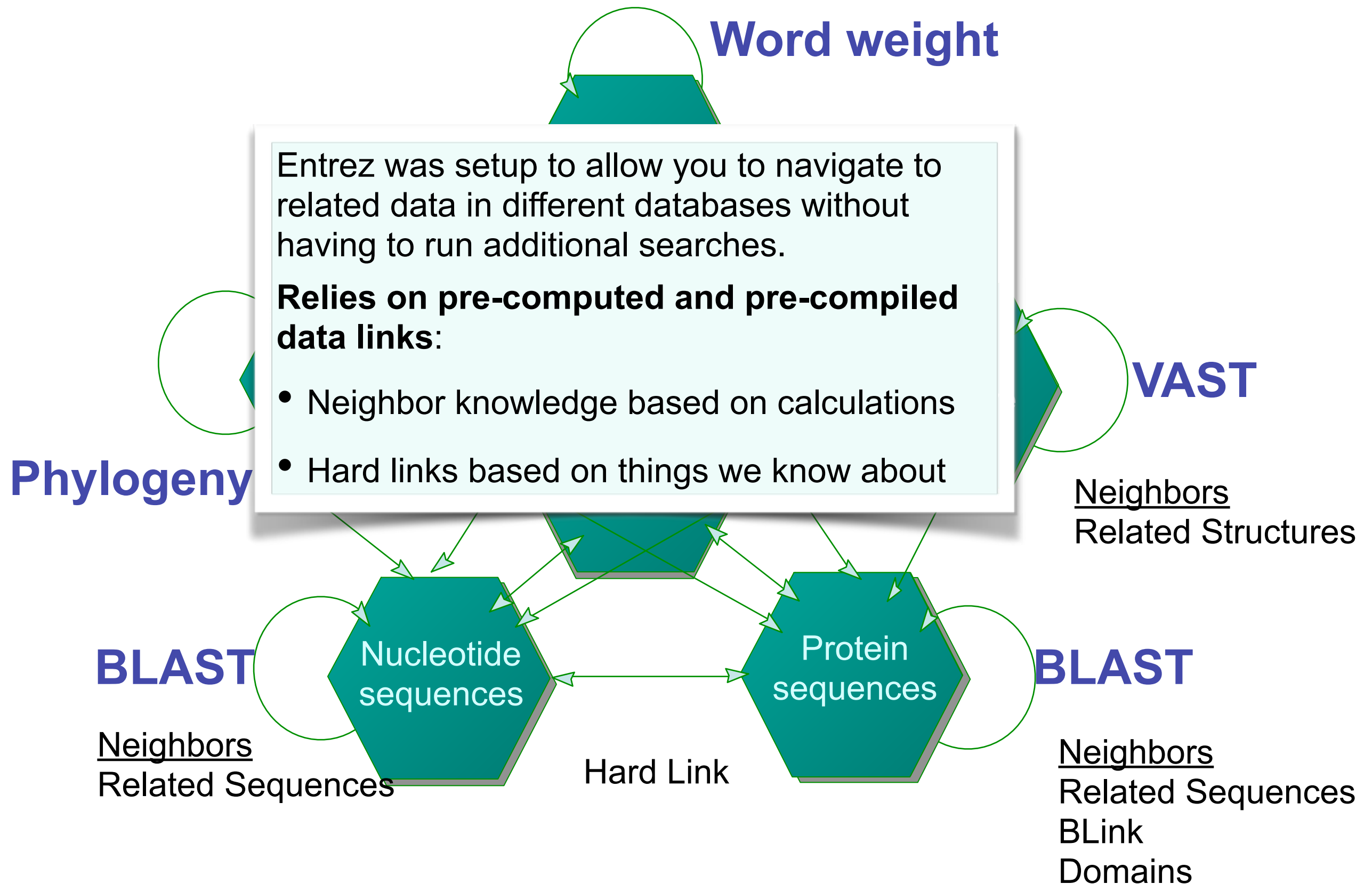
Search

Popular Resources
PubMed
Bookshelf
PubMed Central
PubMed Health
BLAST
Nucleotide
Genome
SNP
Gene
Protein

NCBI Announcements
NCBI's April Newsletter is on the Bookshelf
04 May 2012
Information about May's Discovery Workshop, the new GTR and Assembly
03 May 2012
A Filter Sidebar will be added soon to the PubMed result pages. The useful
30 Apr 2012
Domain Enhanced Lookup Time Accelerated BLAST (DELTA-BLAST)
[More...](#)

Entrez is available from the main NCBI homepage or from the homepage of individual databases

Entrez: navigating across databases



Global Entrez Query: All NCBI Databases

The screenshot displays the NCBI GQuery interface. At the top, the browser address bar shows the URL www.ncbi.nlm.nih.gov/gquery/?term=ras. The NCBI logo and navigation links (Resources, How To) are visible. The main heading is "Search NCBI databases". A search box contains the term "ras", and a "Search" button is to its right. Below the search box, a message states "About 2,978,774 search results for 'ras'".

Two text boxes are overlaid on the screenshot. The first box contains the URL <http://www.ncbi.nlm.nih.gov/gquery/>. The second box contains the text "The Entrez system: 38 (and counting) integrated databases".

The background interface shows a list of databases under the "Literature" and "Health" categories. The "Literature" category includes Books, MeSH, NLM Catalog, PubMed, and PubMed Central. The "Health" category includes ClinVar, dbGaP, and GTR. The "Proteins" category is partially visible at the bottom.

Category	Database	Count	Description
Literature	Books	1,011	books and reports
	MeSH	402	ontology used for PubMed indexing
	NLM Catalog	223	
	PubMed	54,672	
	PubMed Central	96,114	full-text journal articles
Health	ClinVar	759	human variations of clinical significance
	dbGaP	120	genotype/phenotype interaction studies
	GTR	1,879	genetic testing registry
Proteins	EST	3,985	expressed sequence tag sequences
	GEO Profiles	1,622,789	abundance profiles
	HomoloGene	696	homologous gene sets for selected organisms
	PopSet	2,254	sequence sets from phylogenetic and population studies
Proteins	UniGene	4,770	clusters of expressed transcripts

Search Results

Nucleotide [Save search](#) [Limits](#) [Advanced](#) [Help](#)

Display Settings: ☐ Summary, 20 per page, Sorted by Default order

Found 2324 nucleotide sequences. Nucleotide (35) EST (2289)

Results: 1 to 20 of 35 << First < Prev Page 1 of 2 Next > Last >>

☐ [Danio rerio creatine kinase, muscle b \(ckmb\), mRNA](#)
1. 1,463 bp linear mRNA
Accession: NM_001105683.1 GI: 157787180
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio zgc:63663 \(zgc:63663\), mRNA](#)
2. 2,476 bp linear mRNA
Accession: NM_200614.1 GI: 41055386
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, muscle b \(ckmb\), mRNA](#)
3. 1,552 bp linear mRNA
Accession: NM_130932.1 GI: 18858426
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, mitochondrial 2 \(sarcomeric\), mRNA \(cDNA clone MGC:198091 IMAGE:9039080\), complete cds](#)
4. 1,296 bp linear mRNA
Accession: BC171364.1 GI: 213624628
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, mitochondrial 2 \(sarcomeric\), mRNA \(cDNA clone MGC:172259 IMAGE:8798676\), complete cds](#)
5. 1,400 bp linear mRNA
Accession: BC154617.1 GI: 159155933
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

Discovery Column
(sort, filter, link)

Send to: ☐ **Filter your results:**

All (35)
Bacteria (0)
[INSDC \(GenBank\) \(27\)](#)
[mRNA \(32\)](#)
[RefSeq \(8\)](#)
[Manage Filters](#)

Top Organisms [Tree]
Danio rerio (29)
Ictalurus furcatus (6)

Find related data
Database:

Search details
("Danio rerio"[Organism]
OR zebrafish[All Fields])
AND creatine kinase[All
Fields]
 [See more...](#)

Recent activity

Limits

Limits

Published in the last

Any Date

Search Field Tags

Field: All Fields

Source database

Any

Gene Location

Any

Modified in the last

Any Date

Segmented Sequences

Any

Molecule

Any

Exclude

- ☐ STSs
- ☐ working draft
- ☐ TPA
- ☐ patents

Reset

Search

Search Results

Nucleotide [Save search](#) [Limits](#) [Advanced](#) [Help](#)

Display Settings: ☐ Summary, 20 per page, Sorted by Default order

Found 2324 nucleotide sequences. Nucleotide (35) EST (2289)

Results: 1 to 20 of 35 << First < Prev Page 1 of 2 Next > Last >>

☐ [Danio rerio creatine kinase, muscle b \(ckmb\), mRNA](#)
1. 1,463 bp linear mRNA
Accession: NM_001105683.1 GI: 157787180
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio zgc:63663 \(zgc:63663\), mRNA](#)
2. 2,476 bp linear mRNA
Accession: NM_200614.1 GI: 41055386
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, muscle b \(ckmb\), mRNA](#)
3. 1,552 bp linear mRNA
Accession: NM_130932.1 GI: 18858426
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, mitochondrial 2 \(sarcomeric\), mRNA \(cDNA clone MGC:198091 IMAGE:9039080\), complete cds](#)
4. 1,296 bp linear mRNA
Accession: BC171364.1 GI: 213624628
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, mitochondrial 2 \(sarcomeric\), mRNA \(cDNA clone MGC:172259 IMAGE:8798676\), complete cds](#)
5. 1,400 bp linear mRNA
Accession: BC154617.1 GI: 159155933
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

Discovery Column
(sort, filter, link)

Send to: ☐ **Filter your results:**

All (35)
Bacteria (0)
[INSDC \(GenBank\) \(27\)](#)
[mRNA \(32\)](#)
[RefSeq \(8\)](#)
[Manage Filters](#)

Top Organisms [Tree]
Danio rerio (29)
Ictalurus furcatus (6)

Find related data
Database:

Search details
("Danio rerio"[Organism]
OR zebrafish[All Fields])
AND creatine kinase[All
Fields]
 [See more...](#)

Recent activity

Advanced: Search Builder

Nucleotide Advanced Search Builder

zebrafish[Organism] AND "creatine kinase"[Title]

[Clear](#)

Helps build complex fielded queries

Organism [Show index list](#)

AND "creatine kinase"[Title] [Hide index list](#)

[Previous 200](#)

 [Next 200](#)

AND [Refresh index](#) [Show index list](#)

or [Add to history](#)

Items from search history can be included / combined / modified

History

[Clear history](#)

Search	Add to builder	Query	Items found	Time
#7	Add	Search zebrafish[organism] AND actin[title]	71	12:41:16
#4	Add	Search zebrafish actin	1288	12:40:07
#1	Add	Search zebrafish creatine kinase	34	12:39:02

Complex Query Results

Display Settings: ☒ Summary, 20 per page, Sorted by Default order

Send to: ☒ **Filter your results:**

Results: 6

☐ [Danio rerio creatine kinase, brain a \(ckba\), mRNA](#)

1. 1,481 bp linear mRNA
Accession: NM_001077163.1 GI: 116004536
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, mitochondrial 1 \(ckmt1\), nuclear gene encoding mitochondrial protein, mRNA](#)

2. mRNA

☐ [Danio rerio creatine kinase, muscle a \(ckma\), mRNA](#)

3. 1,552 bp linear mRNA
Accession: NM_130932.1 GI: 18858426
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, mitochondrial 2 \(sarcomeric\) \(ckmt2\), nuclear gene encoding mitochondrial protein, mRNA](#)

4. 1,401 bp linear mRNA
Accession: NM_200697.1 GI: 41152341
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, muscle b \(ckmb\), mRNA](#)

5. 1,463 bp linear mRNA
Accession: NM_001105683.1 GI: 157787180
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

☐ [Danio rerio creatine kinase, brain b \(ckbb\), mRNA](#)

6. 1,459 bp linear mRNA
Accession: NM_173222.1 GI: 27545192
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

Filter your results:

All (6)

Bacteria (0)

INSDC (GenBank) (0)

[mRNA \(6\)](#)

[RefSeq \(6\)](#)

[Manage Filters](#)

Analyze these sequences

[Run BLAST](#)

Find related data

Database:

[Find items](#)

Search details

("Danio rerio"[Organism]
AND "creatine kinase"
[Title]) AND "refseq"
[Filter]

[Search](#) [See more...](#)

Recent activity

("Danio rerio"[Organism] AND "creatine kinase"[Title])
AND "refseq"[Filter] AND mrna[Filter]

Controlled Vocabularies

- Taxonomy primary controlled vocabulary / classification system for molecular databases at NCBI

The screenshot shows the NCBI Nucleotide search interface. The search bar contains the text "sponges". Below the search bar are links for "Save search", "Limits", and "Advanced". To the right of the search bar is a "Search" button and a "Help" link. A "Search details" window is open, displaying the search query: "Porifera"[Organism] OR sponges[All Fields].

Nucleotide [Help](#)

[Save search](#) [Limits](#) [Advanced](#)

Search details

```
"Porifera"[Organism]
OR sponges[All Fields]
```

- ▶ Medical Subject Headings (MeSH) primary controlled vocabulary / classification system (ontology) for molecular databases at NCBI

The screenshot shows the PubMed search interface. The search bar contains the text "sponges". Below the search bar are links for "RSS", "Save search", and "Limits". To the left of the search bar is the PubMed logo and the text "US National Library of Medicine National Institutes of Health". To the right of the search bar is a "Search" button and a "Help" link. A "Search details" window is open, displaying the search query: "porifera"[MeSH Terms] OR "porifera"[All Fields] OR "sponges"[All Fields].

PubMed [Help](#)

[RSS](#) [Save search](#) [Limits](#)

Search details

```
"porifera"[MeSH Terms]
OR "porifera"[All
Fields] OR
"sponges"[All Fields]
```


BLAST is a very important tool available from the NCBI Homepage

<http://www.ncbi.nlm.nih.gov/guide/>

The screenshot shows the NCBI homepage with a red border. At the top, there is a navigation bar with "NCBI", "Resources", and "How To" links. Below this is the NCBI logo and a search bar with a dropdown menu set to "All Databases". The main content area is divided into three columns. The left column contains a "NCBI Home" section with a list of links: 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, and Variation. The middle column has a "Welcome to NCBI" section with a description of the center's mission and a "Get Started" section with links to Tools, Downloads, How-To's, and Submissions. Below this is a "PubMed Central" section with a description of its free full-text articles. The right column features a "Popular Resources" section with a list of links: BLAST, Bookshelf, Gene, Genome, Nucleotide, OMIM, Protein, PubChem, PubMed, PubMed Central, and SNP. A red arrow points to the "BLAST" link. Below this is an "NCBI News" section with two news items: "NAR's 2011 Database Issue is out with 9 NCBI-Authorized Papers" and "New articles are available describing the new Epigenomics".

NCBI Resources How To My NCBI

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

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

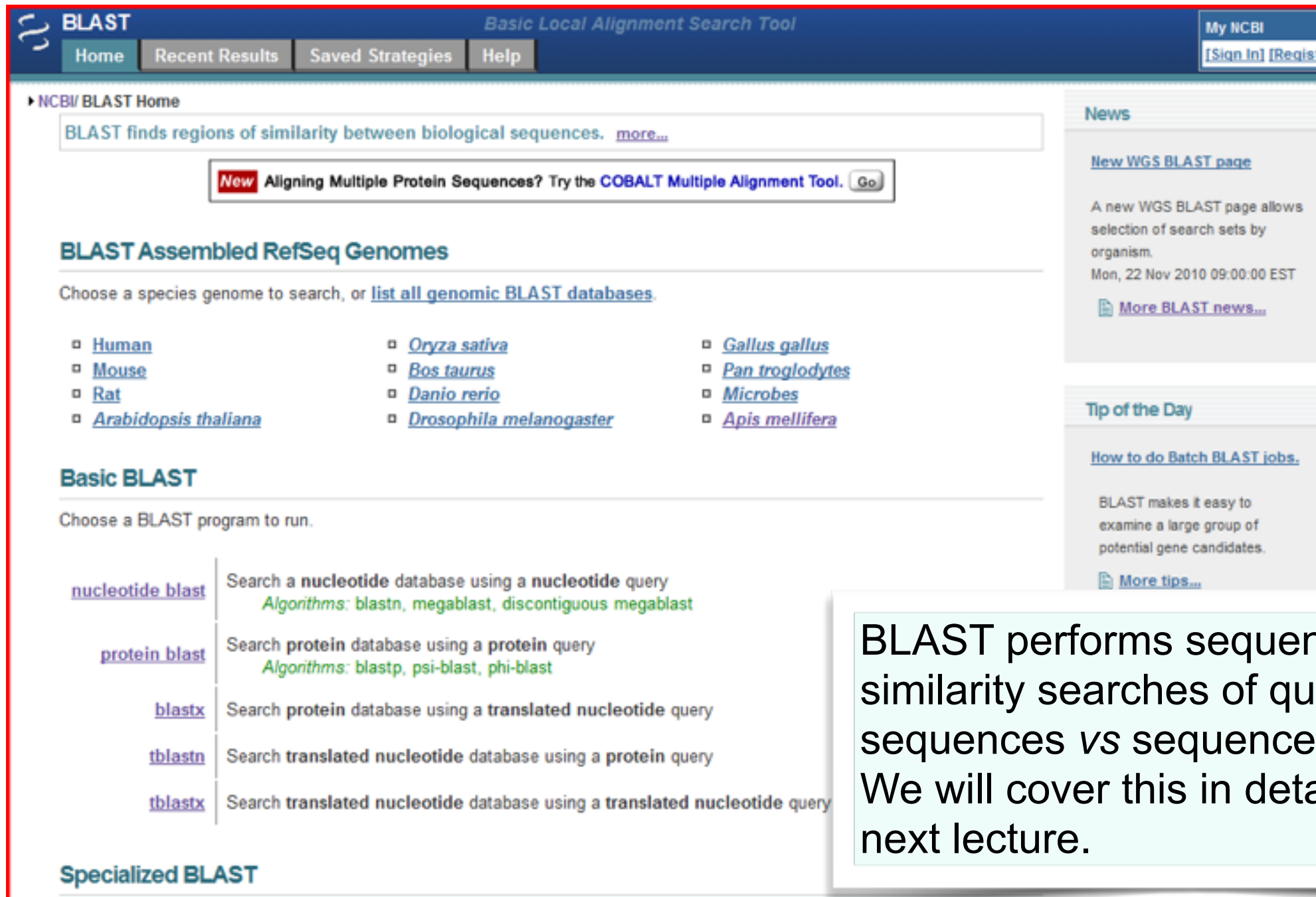
PubMed Central
Free Full Text. Over 1,500,000 articles from over 450 journals. Linked to PubMed and fully searchable.

Popular Resources
BLAST
Bookshelf
Gene
Genome
Nucleotide
OMIM
Protein
PubChem
PubMed
PubMed Central
SNP

NCBI News
NAR's 2011 Database Issue is out with 9 NCBI-Authorized Papers
05 Jan 2011
New articles are available describing the new Epigenomics
New NCBI News Issue
29 Nov 2010

BLAST – Basic Local Alignment Search Tool

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>



The screenshot shows the NCBI BLAST homepage. At the top, there's a navigation bar with 'Home', 'Recent Results', 'Saved Strategies', and 'Help'. A 'My NCBI' section includes links for '[Sign In]' and '[Register]'. Below the navigation bar, a banner states 'BLAST finds regions of similarity between biological sequences.' with a 'more...' link. A 'New' alert box promotes the 'COBALT Multiple Alignment Tool'. The main content area is divided into sections: 'BLAST Assembled RefSeq Genomes' with a list of species (Human, Mouse, Rat, Arabidopsis thaliana, Oryza sativa, Bos taurus, Danio rerio, Drosophila melanogaster, Gallus gallus, Pan troglodytes, Microbes, Apis mellifera), 'Basic BLAST' with a list of programs (nucleotide blast, protein blast, blastx, tblastn, tblastx) and their descriptions, and 'Specialized BLAST'. On the right side, there's a 'News' section with a link to 'New WGS BLAST page' and a 'Tip of the Day' section with a link to 'How to do Batch BLAST jobs.'.

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign In] [Register]

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 <i>Algorithms: blastn, megablast, discontinuous megablast</i>
protein blast	Search protein database using a protein query <i>Algorithms: blastp, psi-blast, phi-blast</i>
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

News

[New WGS BLAST page](#)

A new WGS BLAST page allows selection of search sets by organism.
Mon, 22 Nov 2010 09:00:00 EST

[More BLAST news...](#)

Tip of the Day

[How to do Batch BLAST jobs.](#)

BLAST makes it easy to examine a large group of potential gene candidates.

[More tips...](#)

BLAST performs sequence similarity searches of query sequences vs sequence databases. We will cover this in detail in the next lecture.

SUMMARY

- Bioinformatics is computer aided biology.
- Bioinformatics deals with the collection, archiving, organization, and interpretation of a wide range of biological data.
- There are a large number of primary, secondary and tertiary bioinformatics databases.
- The NCBI and EBI are major online bioinformatics service providers.
- Introduced GenBank, RefSeq, UniProt, PDB databases as well as a number of 'boutique' databases including PFAM and OMIM.
- Introduced the notion of *controlled vocabularies* and *ontologies*.
- Described the use of ENTREZ and BLAST for searching databases.

HOMework

- ☑ Complete the **initial course questionnaire**:
<http://tinyurl.com/bioinf525-questions>
- ☑ Check out the “**Background Reading**” material on Ctools:
<http://tinyurl.com/bioinf525-w16>
- ☑ Complete the **lecture 1.1 homework questions**:
<http://tinyurl.com/bioinf525-quiz1>

THANK YOU

ADDITIONAL DATABASES OF NOTE
(SLIDES FOR YOUR REFERENCE)

NCBI Metadatabases

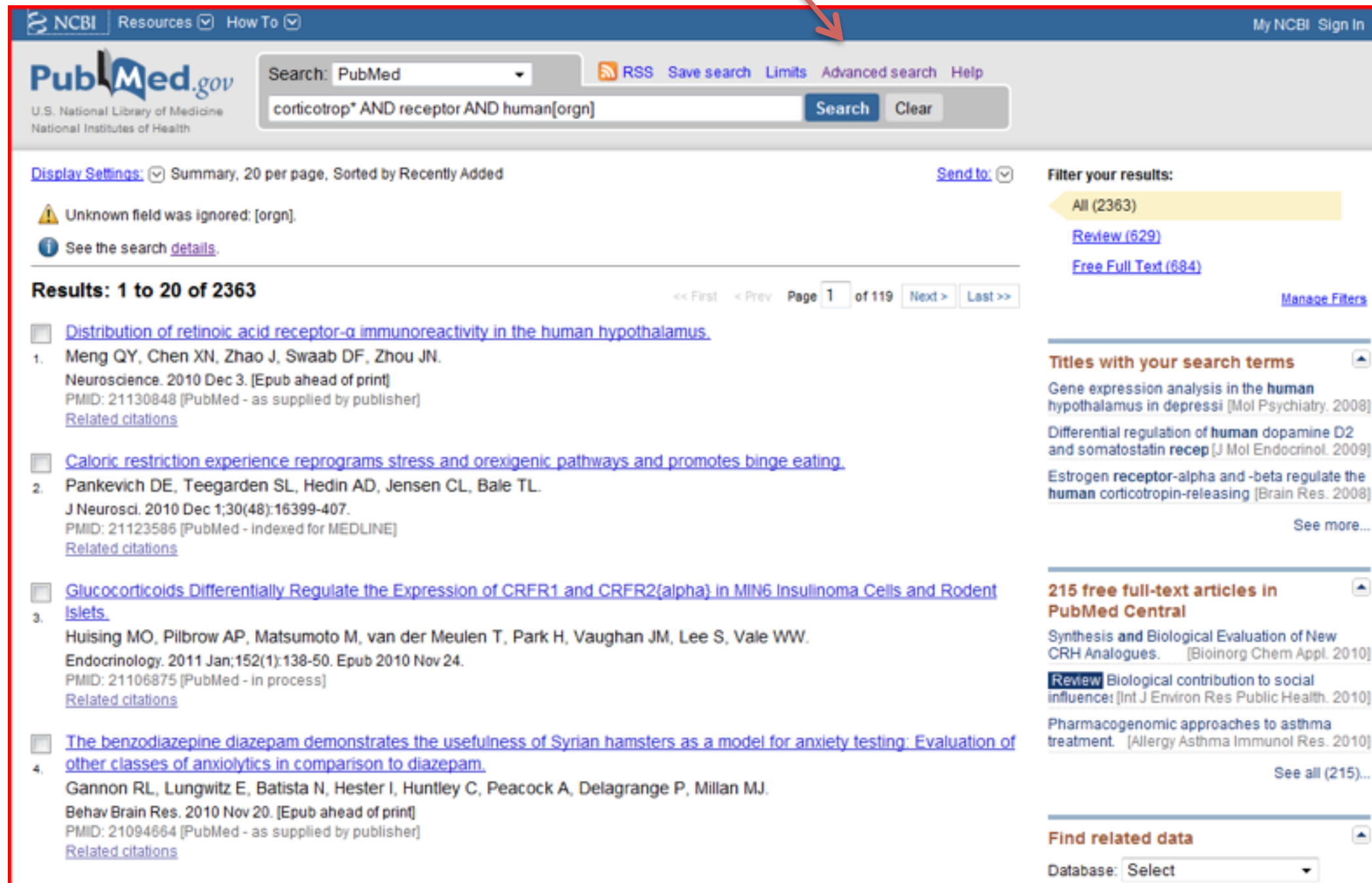
- **Gene**
 - ▶ molecular data and literature related to genes
- **HomoloGene**
 - ▶ automated collection of homologous genes from selected eukaryotes
- **Taxonomy**
 - ▶ access to NCBI data through source organism taxonomic classification
- **PubChem**
 - ▶ small organic molecules and their biological activities
- **BioSystems**
 - ▶ biochemical pathways and processes linked to NCBI genes, gene products, small molecules, and structures

PubMed

- Curated database of biomedical journal articles
- Data records are annotated with MeSH terms (Medical Subject Headings)
- Contract workers actually read all of the articles and classify them with the MeSH terms
- PubMed entries contain article abstracts
- PubMed Central contains full journal articles, but the majority are not freely re-distributable

PubMed results

Limits and Advanced search can be used to refine searches



NCBI Resources How To My NCBI Sign In

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed
corticotrop* AND receptor AND human[orgn] Search Clear

Display Settings: Summary, 20 per page, Sorted by Recently Added
Unknown field was ignored: [orgn].
See the search details.

Results: 1 to 20 of 2363
<< First < Prev Page 1 of 119 Next > Last >>

☐ [Distribution of retinoic acid receptor- \$\alpha\$ immunoreactivity in the human hypothalamus.](#)
1. Meng QY, Chen XN, Zhao J, Swaab DF, Zhou JN.
Neuroscience. 2010 Dec 3. [Epub ahead of print]
PMID: 21130848 [PubMed - as supplied by publisher]
[Related citations](#)

☐ [Caloric restriction experience reprograms stress and orexigenic pathways and promotes binge eating.](#)
2. Pankevich DE, Teegarden SL, Hedin AD, Jensen CL, Bale TL.
J Neurosci. 2010 Dec 1;30(48):16399-407.
PMID: 21123586 [PubMed - Indexed for MEDLINE]
[Related citations](#)

☐ [Glucocorticoids Differentially Regulate the Expression of CRFR1 and CRFR2\[alpha\] in MIN6 Insulinoma Cells and Rodent Islets.](#)
3. Huising MO, Pilbrow AP, Matsumoto M, van der Meulen T, Park H, Vaughan JM, Lee S, Vale WW.
Endocrinology. 2011 Jan;152(1):138-50. Epub 2010 Nov 24.
PMID: 21106875 [PubMed - in process]
[Related citations](#)

☐ [The benzodiazepine diazepam demonstrates the usefulness of Syrian hamsters as a model for anxiety testing: Evaluation of other classes of anxiolytics in comparison to diazepam.](#)
4. Gannon RL, Lungwitz E, Batista N, Hester I, Huntley C, Peacock A, Delagrangre P, Millan MJ.
Behav Brain Res. 2010 Nov 20. [Epub ahead of print]
PMID: 21094864 [PubMed - as supplied by publisher]
[Related citations](#)

Filter your results:
All (2363)
[Review \(629\)](#)
[Free Full Text \(684\)](#)
[Manage Filters](#)

Titles with your search terms
Gene expression analysis in the human hypothalamus in depressi [Mol Psychiatry. 2008]
Differential regulation of human dopamine D2 and somatostatin recep [J Mol Endocrinol. 2009]
Estrogen receptor-alpha and -beta regulate the human corticotropin-releasing [Brain Res. 2008]
[See more...](#)

215 free full-text articles in PubMed Central
Synthesis and Biological Evaluation of New CRH Analogues. [Bioinorg Chem Appl. 2010]
[Review](#) Biological contribution to social influence [Int J Environ Res Public Health. 2010]
Pharmacogenomic approaches to asthma treatment. [Allergy Asthma Immunol Res. 2010]
[See all \(215\)...](#)

Find related data
Database: Select

Small molecule databases have been added at NCBI

<http://pubchem.ncbi.nlm.nih.gov/>

The screenshot displays the PubChem website interface. At the top, a navigation bar includes links for Databases, Deposition, Services, Help, and a 'more' dropdown. The central area features the PubChem logo and three main search categories: BioAssay, Compound, and Substance, each with an icon and a help link. Below these is a search input field with a 'GO' button and a link to 'Advanced search'. A link for 'Chemical structure search | BioActivity analysis' is also present. A news banner highlights the addition of 2.5 million structures from the IBM BAO (Business Analytics and Optimization) strategic IP insight platform (SIIP). A 'more ...' link with a RSS icon is located below the banner. On the right side, a vertical sidebar lists various services: Bioactivity summary, Bioactivity datatable, Bioactivity structure-activity, Chemical structure search, 3D conformer viewer, Chemical structure clustering, Deposition gateway, Structure download, Bioassay download, and PubChem FTP. The footer contains links for 'Write to Helpdesk', 'Disclaimer', 'Privacy Statement', 'Accessibility', and 'Data Citation Guidelines', along with the text 'National Center for Biotechnology Information' and 'NLM | NIH | HHS'.

Databases ▼ Deposition Services ▼ Help more ▼

PubChem

BioAssay ? Compound ? Substance ?

Advanced search

[Chemical structure search](#) | [BioActivity analysis](#)



New More than 2.5 million structures from the **IBM** BAO (Business Analytics and Optimization) strategic IP insight platform (SIIP) are now available in PubChem. See [more..](#) and related news.

[more ...](#)

[Write to Helpdesk](#) | [Disclaimer](#) | [Privacy Statement](#) | [Accessibility](#) | [Data Citation Guidelines](#)
National Center for Biotechnology Information
NLM | NIH | HHS

Bioactivity summary
Bioactivity datatable
Bioactivity structure-activity
Chemical structure search
3D conformer viewer
Chemical structure clustering
Deposition gateway
Structure download
Bioassay download
PubChem FTP

HomoloGene - Homologous genes from different organisms <http://www.ncbi.nlm.nih.gov/homologene>

**HomoloGene**
Discover Homologs

My NCBI
[\[Sign In\]](#) [\[Regis\]](#)

All DatabasesPubMedNucleotideProteinGenomeStructureOMIMPMCJournalsBooks

Search HomoloGene for

LimitsPreview/IndexHistoryClipboardDetails

HomoloGene
Homepage
Query Tips
Build Procedure
FTP site

Genome Resources
[Homo sapiens](#)
[Mus musculus](#)
[Rattus norvegicus](#)
[Danio rerio](#)

HomoloGene is a system for automated detection of homologs among the annotated genes of several completely sequenced eukaryotic genomes.

HomoloGene Release 65 Statistics

Initial numbers of genes from complete genomes, numbers of genes placed in a homology group, and the numbers of groups for each species.

Species	Number of Genes		HomoloGene groups
	Input	Grouped	
Homo sapiens	19,943*	18,981	18,431
Pan troglodytes	25,096	16,850	15,980
Canis familiaris	19,766	16,708	15,951
Bos taurus	22,049	18,180	16,224
Mus musculus	25,388	21,766	19,005
Rattus norvegicus	21,991	19,229	17,473
Gallus gallus	17,959	13,142	11,905
Danio rerio	26,690*	21,084	14,067
Drosophila melanogaster	13,827*	9,282	7,749
Anopheles gambiae	12,460	8,867	7,541
Caenorhabditis elegans	20,132*	8,678	4,810
Schizosaccharomyces pombe	5,043	3,225	2,935
Saccharomyces cerevisiae	5,880	4,851	4,370
Kluyveromyces lactis	5,335	4,459	4,382
Eremothecium gossypii	4,722	3,928	3,884
Magnaporthe grisea	12,832	7,330	6,399
Neurospora crassa	9,821*	6,287	6,144
Arabidopsis thaliana	27,000*	19,961	11,242

What's New

HomoloGene release 65 includes updated annotations for the following species: Homo sapiens (NCBI release 37.2), Danio rerio (NCBI release 4.1), Drosophila melanogaster (NCBI release 9.3) Caenorhabditis elegans (NCBI release 9.1), Arabidopsis thaliana (NCBI release 9.1).

Related Resources

Entrez Genomes
A collection of complete genome sequences that includes more than 1000 viruses and over hundred microbes

- [Archaea](#)
- [Bacteria](#)
- [Eukaryota](#)

Online Mendelian Inheritance in Man – OMIM

<http://www.ncbi.nlm.nih.gov/omim>

The screenshot shows the OMIM website interface. At the top, there is a navigation bar with the NCBI logo, the OMIM title, and a search bar. Below the search bar, there are tabs for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The main content area features a list of search instructions and a welcome message. The left sidebar contains links for 'Entrez', 'OMIM', 'Help', 'FAQ', and 'OMIM Facts'.

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

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

- 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

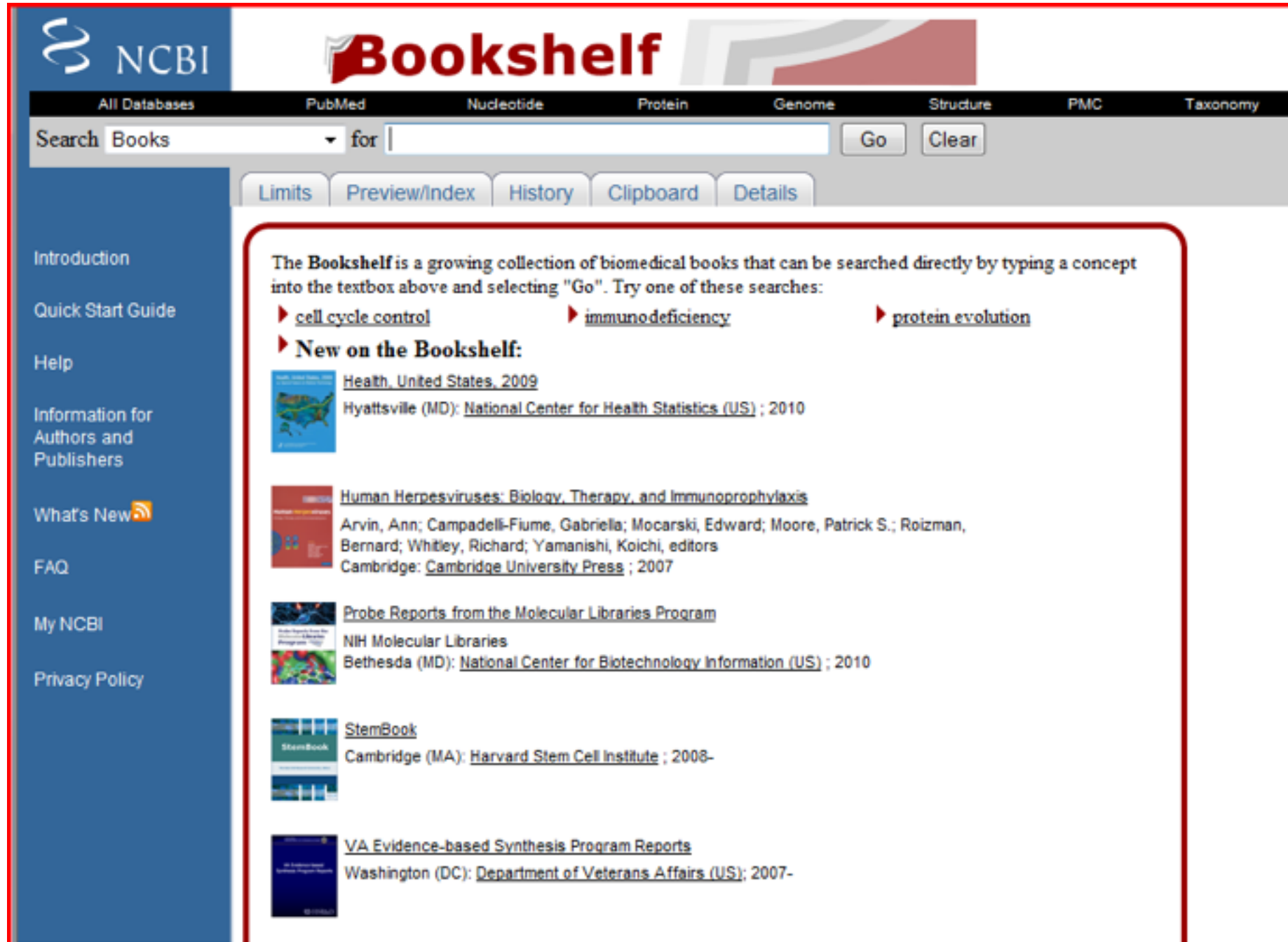
Welcome to OMIM®, Online Mendelian Inheritance in Man®. OMIM is a comprehensive, authoritative, and timely compendium of human genes and genetic phenotypes. The full-text, referenced overviews in OMIM contain information on all known mendelian disorders and over 12,000 genes. OMIM focuses on the relationship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

This database was initiated in the early 1960s by Dr. Victor A. McKusick as a catalog of mendelian traits and disorders, entitled Mendelian Inheritance in Man (MIM). Twelve book editions of MIM were published between 1966 and 1998. The online version, OMIM, was created in 1985 by a collaboration between the National Library of Medicine and the William H. Welch Medical Library at Johns Hopkins. It was made generally available on the

OMIM is essentially a set of reviews of human genes, gene function and phenotypes. Includes causative mutations where known.

The NCBI Bookshelf includes many well known molecular biology texts.

<http://www.ncbi.nlm.nih.gov/books/>



The screenshot shows the NCBI Bookshelf homepage. At the top, the NCBI logo is on the left, and the 'Bookshelf' title is in the center. Below the title is a navigation bar with links to 'All Databases', 'PubMed', 'Nucleotide', 'Protein', 'Genome', 'Structure', 'PMC', and 'Taxonomy'. A search bar is located below the navigation bar, with the text 'Search Books' and a dropdown menu. To the right of the search bar are 'Go' and 'Clear' buttons. Below the search bar is a row of buttons: 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. On the left side of the page is a blue sidebar with links to 'Introduction', 'Quick Start Guide', 'Help', 'Information for Authors and Publishers', 'What's New', 'FAQ', 'My NCBI', and 'Privacy Policy'. The main content area on the right contains a welcome message and a list of new books. The welcome message says: 'The Bookshelf is a growing collection of biomedical books that can be searched directly by typing a concept into the textbox above and selecting "Go". Try one of these searches:'. Below this are three links: 'cell cycle control', 'immunodeficiency', and 'protein evolution'. The 'New on the Bookshelf:' section lists five books with their covers and titles: 'Health, United States, 2009', 'Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis', 'Probe Reports from the Molecular Libraries Program', 'StemBook', and 'VA Evidence-based Synthesis Program Reports'.

NCBI Bookshelf

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy

Search Books for Go Clear



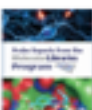


Limits Preview/Index History Clipboard Details

Introduction
Quick Start Guide
Help
Information for Authors and Publishers
What's New
FAQ
My NCBI
Privacy Policy

The **Bookshelf** is a growing collection of biomedical books that can be searched directly by typing a concept into the textbox above and selecting "Go". Try one of these searches:

- ▶ [cell cycle control](#)
- ▶ [immunodeficiency](#)
- ▶ [protein evolution](#)

▶ **New on the Bookshelf:**

-  [Health, United States, 2009](#)
Hyattsville (MD): [National Center for Health Statistics \(US\)](#) ; 2010
-  [Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis](#)
Arvin, Ann; Campadelli-Fiume, Gabriella; Mocarski, Edward; Moore, Patrick S.; Roizman, Bernard; Whitley, Richard; Yamanishi, Koichi, editors
Cambridge: [Cambridge University Press](#) ; 2007
-  [Probe Reports from the Molecular Libraries Program](#)
NIH Molecular Libraries
Bethesda (MD): [National Center for Biotechnology Information \(US\)](#) ; 2010
-  [StemBook](#)
Cambridge (MA): [Harvard Stem Cell Institute](#) ; 2008-
-  [VA Evidence-based Synthesis Program Reports](#)
Washington (DC): [Department of Veterans Affairs \(US\)](#); 2007-

GEO: Gene Expression Omnibus

- Gene expression data (mostly from microarrays but also RNA-seq data, 2 methods for measuring RNA levels)

Query browse and download data sets

The screenshot shows the NCBI GEO website. At the top, there's a navigation bar with the NCBI logo and the GEO logo (Gene Expression Omnibus). Below the navigation bar, there's a description of GEO: "Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. More information »".

The main content area is divided into two sections: "GEO navigation" and "Site contents".

GEO navigation:

- QUERY:** Includes links to "DataSets", "Gene profiles", "GEO accession", and "GEO BLAST". Each of these links has a search input field and a "GO" button.
- BROWSE:** Includes links to "DataSets" and "GEO accessions". "GEO accessions" further branches into "Platforms", "Samples", and "Series".

Site contents:

- Public data:** A table showing the number of public data items:

Category	Count
Platforms	8,246
Samples	514,893
Series	20,827
- Documentation:** Includes links to Overview, FAQ, Find Submission guide, Linking & citing, Journal citations, Construct a Query, Programmatic access, DataSet clusters, GEO announce list, Data disclaimer, and GEO staff.
- Query & Browse:** Includes links to Repository browser and Submitters.

At the bottom left, there is a "Submitter login" link.

- **Series** - (GSExxx) is an original submitter-supplied record that summarizes a study. May contain multiple individual **Samples** (GSMxxx).

Platforms (1) [GPL4091](#) Agilent-014693 Human Genome CGH Microarray 244A (Feature number version)

Samples (4) [GSM495808](#) Aspc1 Cell Line
[GSM495809](#) JH39 Xenograft
[GSM495810](#) JH21 Xenograft

Download family

Download family	Format
SOFT formatted family file(s)	SOFT ?
MINiML formatted family file(s)	MINiML ?
Series Matrix File(s)	TXT ?

Supplementary file	Size	Download	File type/resource
GSE19852_RAW.tar	170.7 Mb	(ftp) (http)	TAR (of TXT)

Raw data provided as supplementary file
 Processed data included within Sample table

[NLM](#) [NIH](#) [GEO Help](#) [Disclaimer](#) [Section 508](#)

Series GSE3541

Status: Public on /

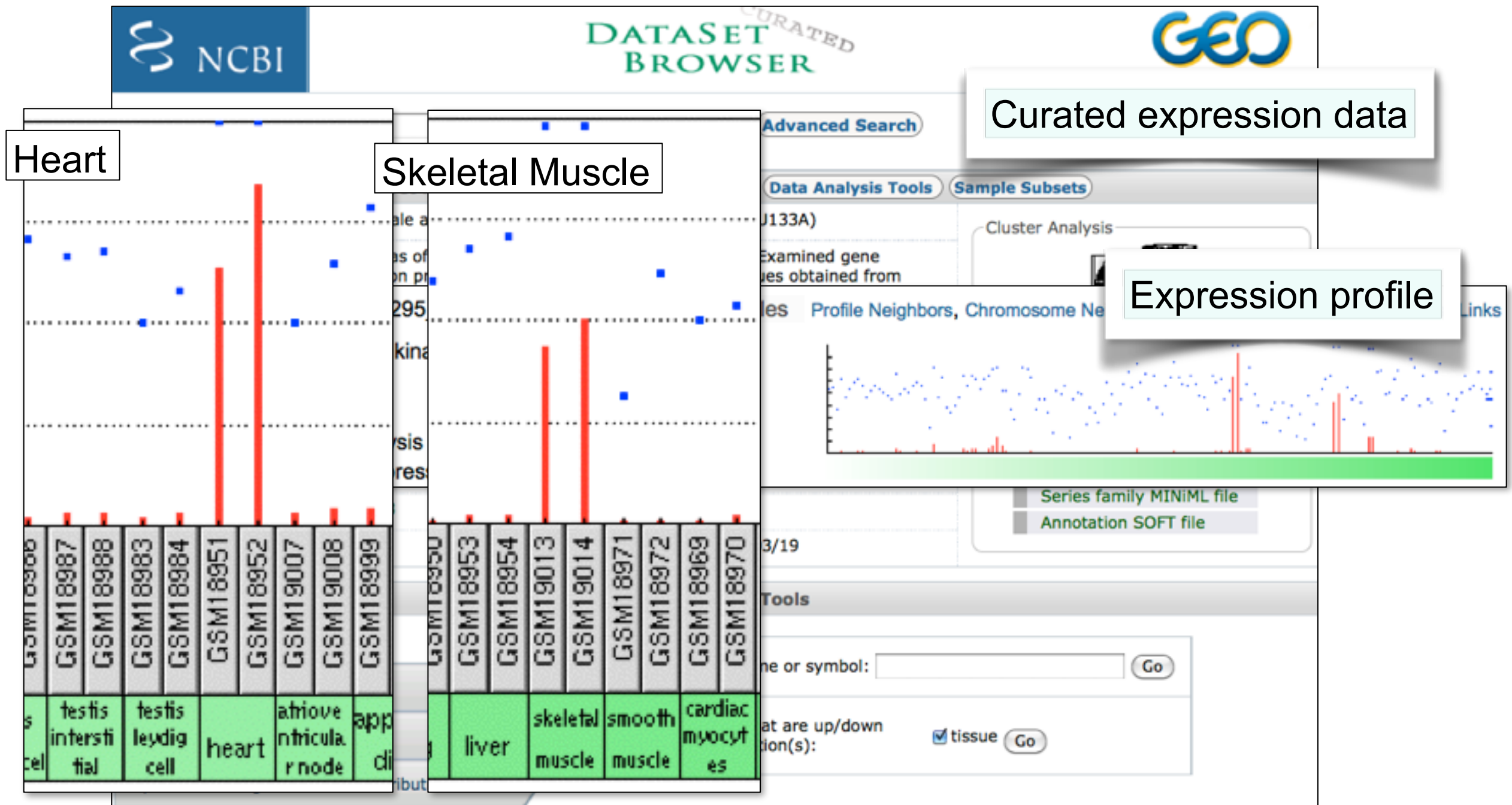
Title: DNA microarray expression data from human gastric cancer cell lines and xenografts

Organism: [Rattus norvegicus](#)

Experiment type: Expression profiling by array

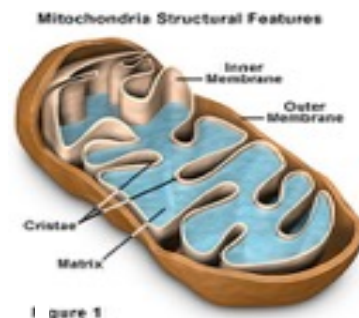
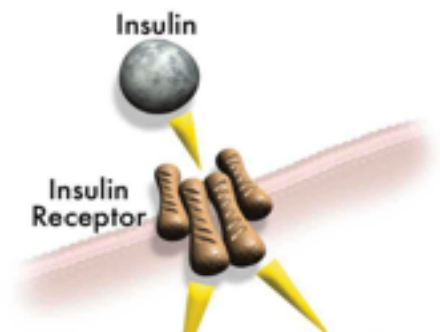
Summary: Mechanical stretching of cells in culture induces a variety of cellular and molecular responses. In this study, we used a microarray to investigate the response of epithelial cells to mechanical stretching. Significance Analysis of Microarrays (SAM) identified 92 genes differentially expressed by strain. Interestingly, several members of the solute carrier family of amino acid transporters, genes involved in amino acid synthesis and development, and amiloride-sensitive epithelial sodium channel gene were induced by strain. These results were confirmed by quantitative real-time polymerase chain reaction (qRT-PCR). Thus, this study identifies genes induced by strain that may be important for amino acid signaling pathways, protein

- **DataSets** - (GDSxxx) are curated collections of selected Samples that are biologically and statistically comparable

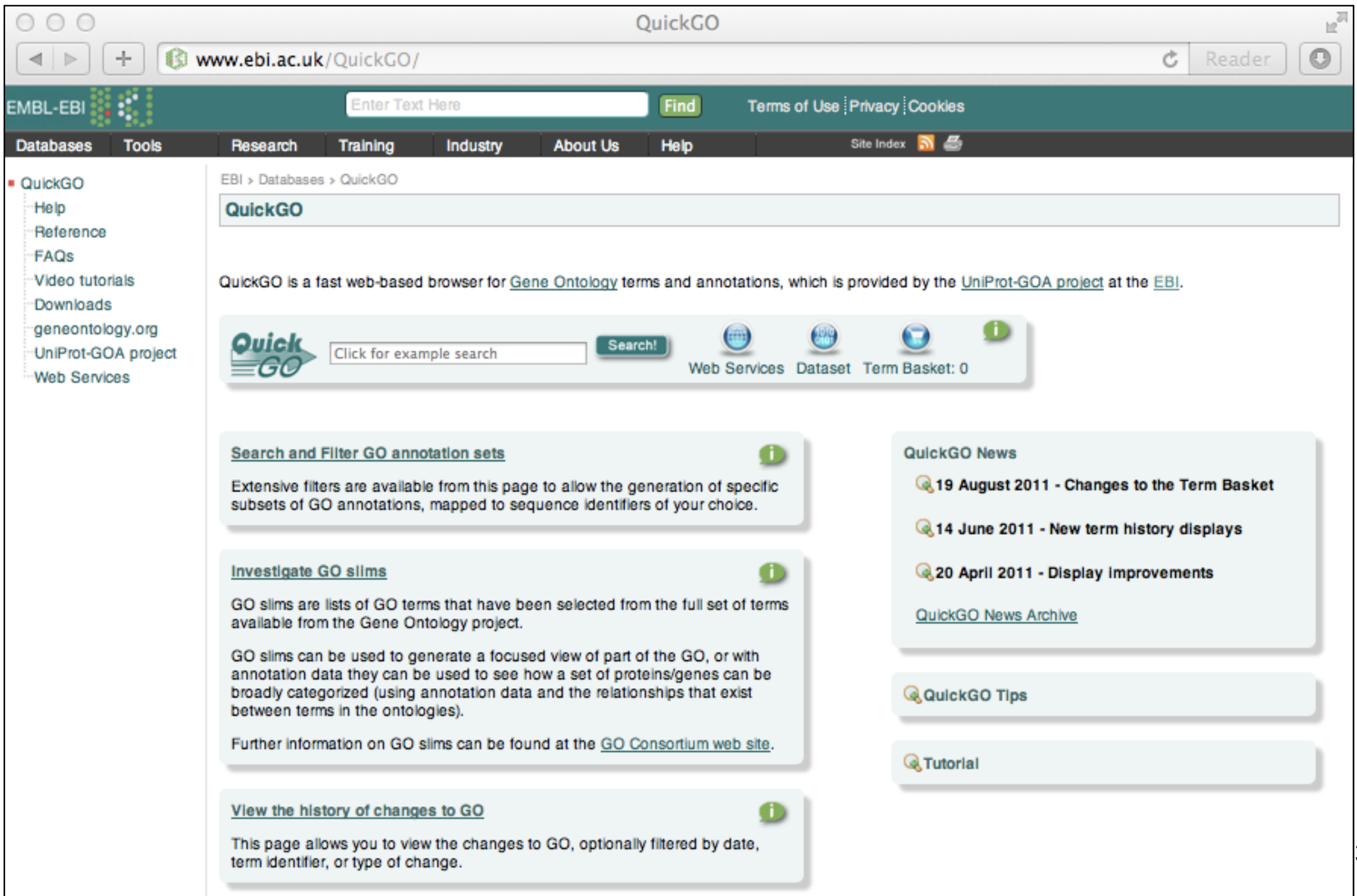


GO Ontologies

- There are three ontologies in GO:
 - ▶ **Biological Process**
A commonly recognized series of events
e.g. cell division, mitosis,
 - ▶ **Molecular Function**
An elemental activity, task or job
e.g. kinase activity, insulin binding
 - ▶ **Cellular Component**
Where a gene product is located
e.g. mitochondrion, mitochondrial
membrane



QuickGO is a fast web-based browser of the Gene Ontology and Gene Ontology annotation data



The screenshot shows the QuickGO web interface in a browser window. The address bar displays www.ebi.ac.uk/QuickGO/. The page features a teal header with the EMBL-EBI logo, a search bar with the placeholder "Enter Text Here", and a "Find" button. Navigation links include "Databases", "Tools", "Research", "Training", "Industry", "About Us", and "Help". A sidebar on the left lists "QuickGO", "Help", "Reference", "FAQs", "Video tutorials", "Downloads", "geneontology.org", "UniProt-GOA project", and "Web Services". The main content area is titled "QuickGO" and includes a description: "QuickGO is a fast web-based browser for [Gene Ontology](#) terms and annotations, which is provided by the [UniProt-GOA project](#) at the [EBI](#)." Below this is a search bar with the placeholder "Click for example search" and a "Search!" button. To the right of the search bar are icons for "Web Services", "Dataset", and "Term Basket: 0". The page is organized into several sections: "Search and Filter GO annotation sets" (with an information icon), "Investigate GO slims" (with an information icon), "View the history of changes to GO" (with an information icon), "QuickGO News" (listing updates from August 2011, June 2011, and April 2011), "QuickGO Tips", and "Tutorial".

QuickGO

EBI > Databases > QuickGO

QuickGO

QuickGO is a fast web-based browser for [Gene Ontology](#) terms and annotations, which is provided by the [UniProt-GOA project](#) at the [EBI](#).

QuickGO

Click for example search

Search!

Web Services Dataset Term Basket: 0

[Search and Filter GO annotation sets](#)

Extensive filters are available from this page to allow the generation of specific subsets of GO annotations, mapped to sequence identifiers of your choice.

[Investigate GO slims](#)

GO slims are lists of GO terms that have been selected from the full set of terms available from the Gene Ontology project.

GO slims can be used to generate a focused view of part of the GO, or with annotation data they can be used to see how a set of proteins/genes can be broadly categorized (using annotation data and the relationships that exist between terms in the ontologies).

Further information on GO slims can be found at the [GO Consortium web site](#).

[View the history of changes to GO](#)

This page allows you to view the changes to GO, optionally filtered by date, term identifier, or type of change.

QuickGO News

- 19 August 2011 - Changes to the Term Basket
- 14 June 2011 - New term history displays
- 20 April 2011 - Display improvements

[QuickGO News Archive](#)

QuickGO Tips

Tutorial

GO annotation in UniProt

An example UniProt entry for hemoglobin beta (HBB_human, P68871) with GO annotation displayed.

The screenshot shows the UniProt entry for Hemoglobin subunit beta (HBB_human, P68871). The browser address bar displays www.uniprot.org/uniprot/P68871. The page title is "Hemoglobin subunit beta - Homo sapiens (Human)". The navigation bar includes links for Names, Attributes, General annotation, Ontologies, Interactions, Sequence annotation, Sequences, References, Web links, Cross-refs, and Entry info. The "Gene Ontology (GO)" section is expanded, showing three categories: Biological_process, Cellular_component, and Molecular_function. Each category lists specific GO terms with their descriptions and sources.

Category	GO Term	Description	Source
Biological_process	bicarbonate transport	Traceable author statement.	Reactome
	blood coagulation	Traceable author statement.	Reactome
	hydrogen peroxide catabolic process	Inferred from direct assay (PubMed 19740759).	BHF-UCL
	nitric oxide transport	Non-traceable author statement (PubMed 8292032).	UniProtKB
	positive regulation of cell death	Inferred from direct assay (PubMed 19740759).	BHF-UCL
	positive regulation of nitric oxide biosynthetic process	Non-traceable author statement (PubMed 7965120).	UniProtKB
	protein heterooligomerization	Inferred from direct assay (PubMed 19740759).	BHF-UCL
	regulation of blood pressure	Inferred from electronic annotation.	UniProtKB-KW
	regulation of blood vessel size	Inferred from electronic annotation.	UniProtKB-KW
	renal absorption	Inferred from mutant phenotype (PubMed 18465053, PubMed 18974585).	UniProtKB
	small molecule metabolic process	Traceable author statement.	Reactome
	Cellular_component	endocytic vesicle lumen	Traceable author statement.
extracellular region		Traceable author statement.	Reactome
haptoglobin-hemoglobin complex		Inferred from direct assay (PubMed 19740759).	BHF-UCL
hemoglobin complex		Non-traceable author statement (Ref.33, Ref.72).	UniProtKB
Molecular_function	heme binding	Inferred from electronic annotation.	InterPro

GO annotation in UniProt

An example UniProt entry for hemoglobin beta (HBB_human, P68871) with GO annotation displayed.

The screenshot displays a web browser window showing the UniProt entry for Hemoglobin subunit beta (HBB_human, P68871). The browser's address bar shows the URL www.uniprot.org/uniprot/P68871. The page title is "Hemoglobin subunit beta - Homo sapiens (Human)".

Below the UniProt header, a QuickGO window is open, showing the GO term "GO:0020037 heme binding". The QuickGO window has a search bar with the text "Click for example search" and a "Search!" button. The QuickGO logo is also visible.

The QuickGO window displays the following information for the term "GO:0020037 heme binding":

- ID:** GO:0020037
- Name:** heme binding
- Ontology:** Molecular Function
- Definition:** Interacting selectively and non-covalently with heme, any compound of iron complexed in a porphyrin (tetrapyrrole) ring.
- GONUTS:** [GO:0020037 Wiki Page](#)

Below the definition, there are tabs for "Synonyms", "Annotation Guidance", "Cross-Ontology Relations", and "Cross-references". The "Synonyms" tab is selected, showing a table of synonyms:

Type	Synonym
exact	haem binding

At the bottom of the QuickGO window, there is a note: "Synonyms are alternative words or phrases closely related in meaning to the term name, with indication of the relationship between the name and synonym given by the synonym scope. Click on the icon for more details." The footer of the QuickGO window contains the text: "Please send comments, suggestions or bug reports to goa@ebi.ac.uk. Click [here](#) for details of how to cite UniProt-GOA and QuickGO."

DAVID: a online tool for assessing GO term enrichment in gene lists

The screenshot shows the DAVID Bioinformatics Resources 6.7 website. The browser address bar displays david.abcc.ncifcrf.gov/home.jsp. The page header includes the DAVID logo and the text "DAVID Bioinformatics Resources 6.7 National Institute of Allergy and Infectious Diseases (NIAID), NIH". A navigation menu at the top contains links: Home, Start Analysis, Shortcut to DAVID Tools, Technical Center, Downloads & APIs, Term of Service, Why DAVID?, and About Us. A left sidebar lists various tools: Functional Annotation (with sub-links for Clustering, Chart, and Table), Gene Functional Classification, Gene ID Conversion, Gene Name Batch Viewer, and NIAID Pathogen Annotation Browser. The main content area features a search bar, a "What's Important in DAVID?" section with links to the release note, citation requirements, and supported arrays, and a list of features with checkboxes: Identify enriched terms, Discover enriched terms, Cluster redundant terms, Visualize genes, Display related terms, Search for other functionally related genes not in the list, List interacting proteins, and Explore gene names in batch. A text box in the lower right corner states: "DAVID allows you to upload lists of genes and search for enriched GO and search for functionally related genes not in your list" with the URL <http://david.abcc.ncifcrf.gov>.

DAVID Functional Annotation Bioinformatics Microarray Analysis

david.abcc.ncifcrf.gov/home.jsp

DAVID Functional Annotation Bioinformatics Microarray Analysis

DAVID: Database for Annotation, Visualization, and Integrated...

DAVID Bioinformatics Resources 6.7
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us

Shortcut to DAVID Tools

- Functional Annotation
 - Functional Annotation Clustering
 - Functional Annotation Chart
 - Functional Annotation Table
- Gene Functional Classification
- Gene ID Conversion
- Gene Name Batch Viewer
- NIAID Pathogen Annotation Browser

Gene-annotation enrichment analysis, functional annotation clustering, KEGG pathway mapping, gene association, homologue match, literature match and [more](#)

Gene Functional Classification

Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. [More](#)

Gene ID Conversion

Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. [More](#)

Gene Name Batch Viewer

Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. [More](#)

published in *Nature Protocols* describes step-by-step procedure to use DAVID!

ne to DAVID 6.7

003 - 2013

What's Important in DAVID?

- [Current \(v 6.7\) release note](#)
- [New requirement to cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supported](#)

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an [update to the sixth version](#) of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand a list of genes. For any list of genes. For any to:

- ☒ Identify enriched terms
- ☒ Discover enriched terms
- ☒ Cluster redundant terms
- ☒ Visualize genes
- ☒ Display related terms
- ☒ Search for other functionally related genes not in the list
- ☒ List interacting proteins
- ☒ Explore gene names in batch

DAVID allows you to upload lists of genes and search for enriched GO and search for functionally related genes not in your list

<http://david.abcc.ncifcrf.gov>

Example output: enriched functions from GO

DAVID: Database for Annotation, Visualization, and Integrat...ID); Science Applications International Corporation (SAIC)

david.abcc.ncifcrf.gov/chartReport.jsp?annot=25

DAVID: Functional Annotation Result Summary

DAVID Bioinformatics Resources 6.7
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Functional Annotation Chart

[Help and Manual](#)

Current Gene List: List_1
Current Background: Homo sapiens
14 DAVID IDs

Options

Rerun Using Options Create Sublist

10 chart records [Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_FAT	regulation of osteoclast differentiation	RT		2	14.3	2.1E-2	9.9E-1
<input type="checkbox"/>	GOTERM_BP_FAT	response to organic substance	RT		4	28.6	2.9E-2	9.6E-1
<input type="checkbox"/>	GOTERM_BP_FAT	regulation of myeloid leukocyte differentiation	RT		2	14.3	3.9E-2	9.5E-1
<input type="checkbox"/>	GOTERM_BP_FAT	positive regulation of transcription from RNA polymerase II promoter	RT		3	21.4	4.8E-2	9.4E-1
<input type="checkbox"/>	GOTERM_BP_FAT	regulation of myeloid cell differentiation	RT		2	14.3	6.5E-2	9.5E-1
<input type="checkbox"/>	GOTERM_BP_FAT	cartilage development	RT		2	14.3	6.9E-2	9.3E-1
<input type="checkbox"/>	GOTERM_BP_FAT	positive regulation of transcription, DNA-dependent	RT		3	21.4	7.5E-2	9.2E-1
<input type="checkbox"/>	GOTERM_BP_FAT	positive regulation of RNA metabolic process	RT		3	21.4	7.6E-2	8.9E-1
<input type="checkbox"/>	GOTERM_BP_FAT	response to protein stimulus	RT		2	14.3	9.8E-2	9.3E-1
<input type="checkbox"/>	GOTERM_BP_FAT	positive regulation of transcription	RT		3	21.4	1.0E-1	9.1E-1

8 gene(s) from your list are not in the output.

Please cite *Nature Protocols* 2009; 4(1):44 & *Genome Biology* 2003; 4(5):P3 within any publication that makes use of any methods inspired by DAVID.