

# Introduction to Internet Resources

成功大學分子醫學研究所 陳嘉興

# Useful Bio-info on Internet

- **Tutorials**
- **Research Tools**
  - lab protocols
  - bioinformatics tools
  - instrument information
- **University/College/Institute**
- **Discussion Groups**
- **Information: citations, museums, databases**

# Internet Resources

## Important genome centers

- National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/>)
- European Bioinformatics Institute (EBI ; <http://www.ebi.ac.uk/>)
- The J. Craig Venter Institute (<http://www.jcvi.org/>)
- The PIGBASE (<http://www.animalgenome.org/pig/maps/index.html>)
- Database Of Genome Sizes (DOGS; <http://www.cbs.dtu.dk/databases/DOGS/index.php>)

# Internet Resources

## Genome Browsers

- **NCBI Map Viewer**
- **UCSC Genome Browser**
- **Ensembl Genome Browser**

# Genome Browser: NCBI

The screenshot shows the NCBI Map Viewer interface. The browser window title is "Map Viewer - Windows Internet Explorer" and the address bar shows "http://www.ncbi.nlm.nih.gov/mapview/". The page content includes a search box, a "Tools Legend" section, and a main area with a hierarchical list of taxonomic groups. The "Tools Legend" includes: Search or Browse the Genome (Q), BLAST (B), Clone Finder (C), Go to region on a chromosome (R), and Genome Resources page (G). The main area lists groups such as Vertebrates (17), Mammals (14), Primates (3), Rodents (2), and Fungi (17). Each group has a table of genome builds and associated tools.

Scientific name	Common name	Build	Tools
<i>Homo sapiens</i>	human	<a href="#">Build 37.1</a>	Q B R C G
		<a href="#">Build 36.3</a>	Q B R C
<i>Macaca mulatta</i>	rhesus macaque	<a href="#">Build 1.1</a>	Q B R G
<i>Pan troglodytes</i>	chimpanzee	<a href="#">Build 2.1</a>	Q B R G
<i>Mus musculus</i>	laboratory mouse	<a href="#">Build 37.1</a>	Q B R C G
		<a href="#">Build 36.1</a>	Q B R
<i>Rattus norvegicus</i>	rat	<a href="#">RGSC v3.4</a>	Q B R G
<i>Aspergillus clavatus</i>		<a href="#">Build 1.1</a>	Q B G
<i>Aspergillus fumigatus</i>		<a href="#">Build 2.1</a>	Q B R G

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

# Genome Browser: UCSC

The screenshot shows the UCSC Genome Browser Gateway interface. The browser window title is "Human (Homo sapiens) Genome Browser Gateway - Windows Internet Explorer". The address bar shows the URL "http://genome.ucsc.edu/cgi-bin/hgGateway". The page has a blue navigation bar with links: Home, Genomes, Blat, Tables, Gene Sorter, PCR, Session, FAQ, Help. Below the navigation bar is the title "Human (Homo sapiens) Genome Browser Gateway" and a paragraph: "The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#). Software Copyright (c) The Regents of the University of California. All rights reserved."

The main search area contains the following form:

clade	genome	assembly	position or search term	gene	image width	
Mammal	Human	Feb. 2009 (GRCh37/hg19)	chr21:33,031,597-33,041,570		800	submit

Below the form, there is a link: "Click here to reset the browser user interface settings to their defaults." and three buttons: "add custom tracks", "configure tracks and display", and "clear position".

The page also features a section titled "About the Human Feb. 2009 (GRCh37/hg19) assembly ([sequences](#))" with the text: "The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#)." Below this is a section "Sample position queries" with the text: "A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information."

At the bottom of the page, there is a table with two columns: "Request:" and "Genome Browser Response:". The first row shows "chr7" and "Display all of chromosome 7".

The browser window shows the Windows taskbar at the bottom with the Start button, several open applications (Microsoft PowerPoint, Human (Homo sapiens...)), and the system tray showing the time as 下午 05:03.

<http://genome.ucsc.edu/cgi-bin/hgGateway>

# Genome Browser: Ensembl

The screenshot shows the Ensembl Genome Browser website in a Windows Internet Explorer browser window. The address bar displays <http://www.ensembl.org/index.html>. The browser's taskbar at the bottom shows several open applications: Microsoft PowerPoint, Ensembl Genome Browser, and PowerPoint 投影片放... The system tray on the right indicates the time as 下午 05:05.

The website interface includes a search bar with a dropdown menu set to "All species" and a "Go" button. Below the search bar, there is a section titled "Browse a Genome" with a description: "The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online. Click on a link below to go to the species' home page." Underneath, there are links for "Popular genomes" with a note to "Log in to customize this list". Three species are listed with icons: Human (GRCh37), Mouse (NCBIM37), and Zebrafish (Zv8). Below this is a section for "All genomes" with a dropdown menu set to "-- Select a species --" and a link to "View full list of all Ensembl species".

On the right side of the page, there is a "New to Ensembl?" section with a list of links: "Learn how to use Ensembl" (with video tutorials and walk-throughs), "Add custom tracks" (using the new Control Panel), "Upload and analyse your data" (and save it to your Ensembl account), "Search for a DNA or protein sequence" (using BLAST or BLAT), "Fetch only the data you want" (from the public database, using the Perl API), "Download our databases via FTP" (in FASTA, MySQL and other formats), and "Mine Ensembl with BioMart" (and export sequences or tables in text, html, or Excel format). A "Did you know...?" box highlights a "NEW!" preliminary assembly of the common baboon (*Papio hamadryas*) available on the pre-ensembl site, with the URL <http://pre.ensembl.org/Baboon>. At the bottom of the right column, there is a "What's New in Release 58 (27 May 2010)" section listing "WormBase WS210 (*C.elegans*)" and "Zebrafish RNASeq models (Zebrafish)".

<http://www.ensembl.org/>

# Internet Resources

## Types of Databases

- **Original / Databank**
- **Experimental (observational) / Computational (predictive) / Curation / Literature**
- **Specialized / Integrated**
- **Public / Commercial / Private**
- **Text / Images / Movies**



# Internet Resources

## Databases

- **References**
- **Sequences**
  - DNA, cDNA, splicing, SNP
  - Protein, domains, structure
- **Others**
  - Promoter
  - Pathway
  - UTRs
  - .....

# Reference Database



<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>

# Reference Database

The screenshot shows a Windows Internet Explorer browser window displaying the NCBI Bookshelf website. The address bar shows the URL <http://www.ncbi.nlm.nih.gov/sites/entrez?db=books>. The page features a search bar with the word "Books" selected, a "Search" button, and links for "Limits" and "Advanced". Below the search bar is a large banner with a green microscopic image and the text: "Bookshelf provides free access to books and documents in life science and healthcare. A vital node in the data-rich resource network at NCBI, Bookshelf enables users to easily browse, retrieve, and read content, and spurs discovery of related information." The page is organized into three columns: "Getting Started" with links for "Quick Start Guide", "Help", "FAQ", and "Tutorials"; "Read" with links for "Browse Titles" and "RSS Feed"; and "More Information" with links for "NLM LitArch (NLM Literature Archive)", "Authors and Publishers", "Librarians", and "Copyright and Permissions". The bottom of the browser window shows the Windows taskbar with various application icons and the system clock displaying "下午 01:47".

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books>

# Reference Database

OMIM Home - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim

NCBI OMIM Online Mendelian Inheritance in Man Johns Hopkins University

All Databases PubMed Nucleotide Protein Genome Structure PMC OMIM

Search OMIM for  Go Clear

Limits Preview/Index History Clipboard Details

- 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 internet starting in 1987. In 1995, OMIM was developed for the World Wide Web by NCBI, the National Center for Biotechnology Information.

OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh.

NLM's Profiles in Science -- The McKusick Papers [More...](#)

NOTE: OMIM is intended for use primarily by physicians and other professionals concerned with genetic disorders, by genetics researchers, and by advanced students in science and medicine. While the OMIM database is open to the public, users seeking information...

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>

# Reference Database



<http://scholar.google.com.tw/>

# Reference Database

JoVE | Journal of Visualized Experiments - Windows Internet Explorer

http://www.jove.com/

promoter prediction

Search

Subscribe About Editorial Boards Authors Blog

Welcome. You do not have subscription access to any sections. [Learn More](#)

**jove** Separate search terms with commas Search Browse All Sign In

Introducing JoVE **AP Applied Physics**

Articles in JoVE Applied Physics will be freely available until September 1, 2012

The *Journal of Visualized Experiments* (JoVE) is a peer reviewed, PubMed-indexed video journal. Our mission is to increase the productivity of scientific research.

1831 Video Articles Published	50 Video Articles per Month	5758 Published Authors
-------------------------------------	-----------------------------------	------------------------------

Publish in JoVE Recommend to Librarian

This month in JoVE July, 2012

3 of 8 Watch this video now

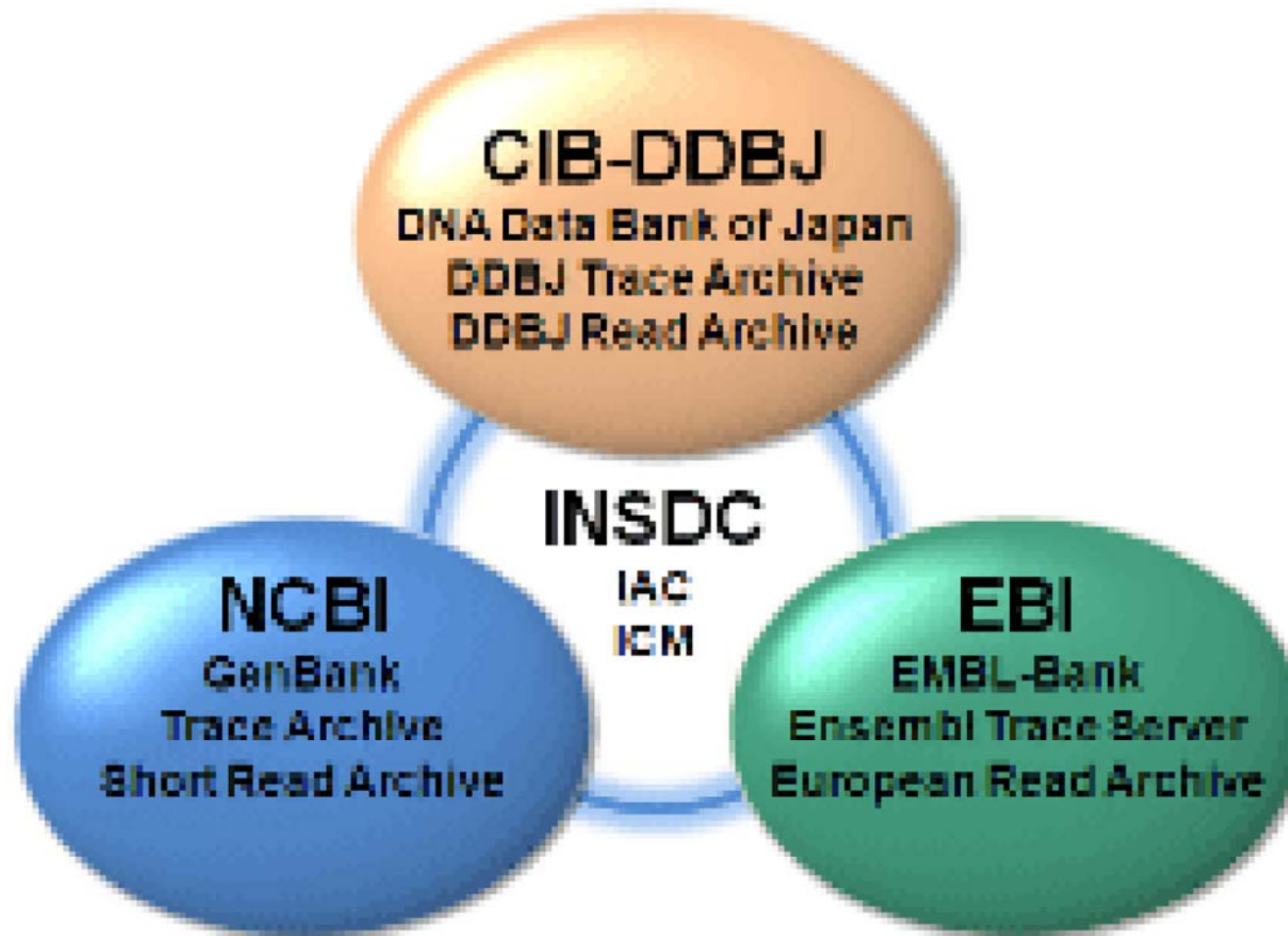
Fabrication of Silica Ultra High Quality Factor ... Published Today

完成 網際網路 | 受保護模式: 啟動 100% 上午 06:10

<http://www.jove.com/>

# Nucleotide Database

## International DNA data banks



# Nucleotide Database: GenBank

GenBank Home - Windows Internet Explorer  
http://www.ncbi.nlm.nih.gov/genbank/

GenBank Nucleotide Search

GenBank Overview

**What is GenBank?**

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

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

An example of a GenBank [record](#) may be viewed for a *Saccharomyces cerevisiae* gene.

**Access to GenBank**

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with [Entrez Nucleotide](#), which is divided into three divisions: [CoreNucleotide](#) (the main collection), [dbEST](#) (Expressed Sequence Tags), and [dbGSS](#) (Genome Survey Sequences).
- Search and align GenBank sequences to a query sequence using [BLAST](#) (Basic Local Alignment Search Tool). BLAST searches [CoreNucleotide](#), [dbEST](#), and [dbGSS](#) independently; see [BLAST info](#) for more information about the numerous BLAST databases.

**GenBank Resources**

- [GenBank Home](#)
- [Submission Types](#)
- [Submission Tools](#)
- [Search GenBank](#)
- [Update GenBank Records](#)

下午 01:52

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



# Nucleotide Database: EMBL

The screenshot shows the EMBL-EBI Databases website. The browser window title is "Bioinformatics Databases | EBI - Windows Internet Explorer". The address bar shows "http://www.ebi.ac.uk/Databases/". The page features a search bar with "All Databases" selected and a search input field. A navigation menu includes "Databases", "Tools", "EBI Groups", "Training", "Industry", "About Us", and "Help". The main content area is titled "Databases at the EBI" and describes the EBI's mission. It lists several databases managed by EBI:

- [ENA](#) - Europe's primary comprehensive nucleotide sequence data resource.
- [UniProt Knowledgebase](#) - a complete annotated protein sequence database.
- [Protein Databank in Europe Database](#) - European Project for the management and distribution of data on macromolecular structures.
- [ArrayExpress](#) - for gene expression data
- [Ensembl](#) - Providing up to date completed metazoic genomes and the best possible automatic annotation.
- [IntAct](#) - Provides a freely available, open source database system and analysis tools for protein interaction data.
- [Databases A-Z](#) - A complete listing of all the EBI databases.

Below this list, it mentions other databases like [Medline](#) and provides a link to browse the available databases. There are also sections for "Patent Data Resources" and "UniProt".

<http://www.ebi.ac.uk/Databases/>

# Nucleotide Database: DDBJ

DDBJ:DNA Data Bank of Japan

Accession DNA Protein AlIBDs Taxonomy Site Search

Accession numbers  Go

DDBJ  UniProt  PDB  DAD  PRF  Patent >>more

HOME Submission How to Use Search/Analysis FTP/WebAPI Report/Statistics Contact Us RSS Japanese

About DDBJ  
How to Use  
Q and A

Sequence Submission

- SAKURA
- Mass Submission
- Data Updates
- DDBJ Read Archive
- DDBJ Trace Archive

Search

- getentry
- ARSA
- TxSearch
- BLAST
- PSI-BLAST
- FASTA
- SSEARCH

Phylogenetics

- ClustalW

Genome Analysis

- GIB
- GIB-V

DDBJ : DNA Data Bank of Japan

DDBJ (DNA Data Bank of Japan) is one of the three databanks that constitute DDBJ/EMBL/GenBank International Nucleotide Sequence Database, which was established through cooperative work with EBI in Europe and NCBI in the USA. Photo: S. Nagira

Hot Topics

- May. 24, 2010 [Modification of way of getting the large size search result by getentry](#)
- May. 14, 2010 [Started Twitter](#)
- May. 14, 2010 [DDBJ/EMBL/GenBank Feature Table Definition revised](#)

Maintenance

- Apr. 21, 2010 [Suspension of some DDBJ activities in Japanese holidays\(4/29,5/1-5\)](#)
- Mar. 16, 2010 [\(Apr.23\)ARSA database search \(DDBJ, DAD\) temporary unavailable](#)
- Feb. 03, 2010 [\(Important!\)Termination of a part of DDBJ services](#)

Sequence Data Submission

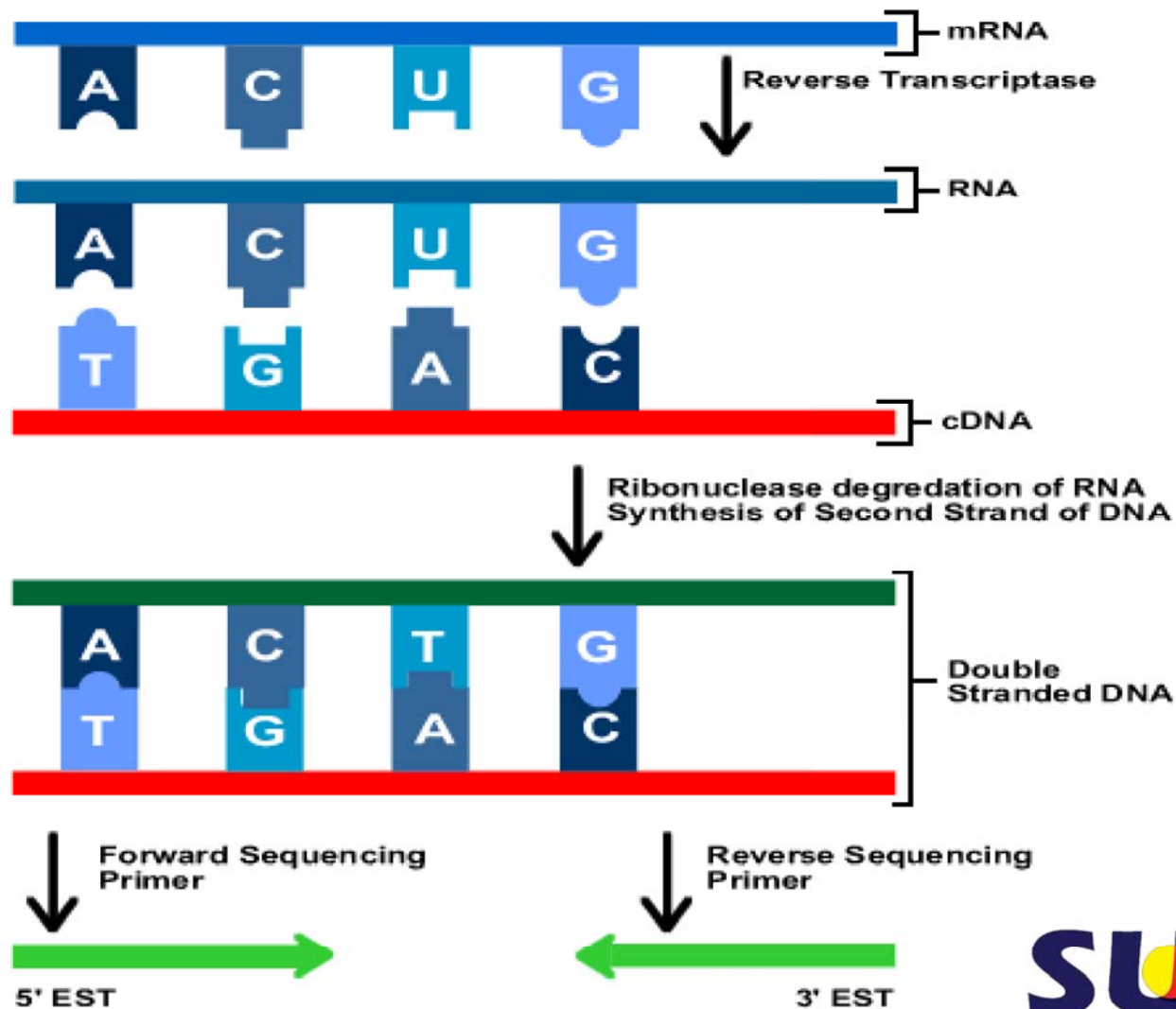
- Submit my sequences**  
Orientation for the data submission
- Update my entries**  
Guidance for the update of the entry

FTP/Web API

- FTP ( ftp.ddbj.nig.ac.jp )**  
Download data files
- Web API**  
Programmatic interfaces of DDBJ Web services

<http://www.ddbj.nig.ac.jp/>

# Nucleotide Database Expression Sequence Tag



# mRNA/cDNA Database

The screenshot shows a Windows Internet Explorer browser window displaying the NCBI dbEST website. The address bar shows the URL <http://www.ncbi.nlm.nih.gov/dbEST/index.html>. The browser's taskbar at the bottom shows several open applications, including Microsoft PowerPoint and the current browser window.

The website content includes the NCBI logo and the title "Expressed Sequence Tags database". Below the title is a navigation menu with links to PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. A search bar is present with the text "Search EST for" and a "Go" button. Below the search bar, there is a dropdown menu for "modified during the last" with "10 Years" selected.

The main content area features several sections:

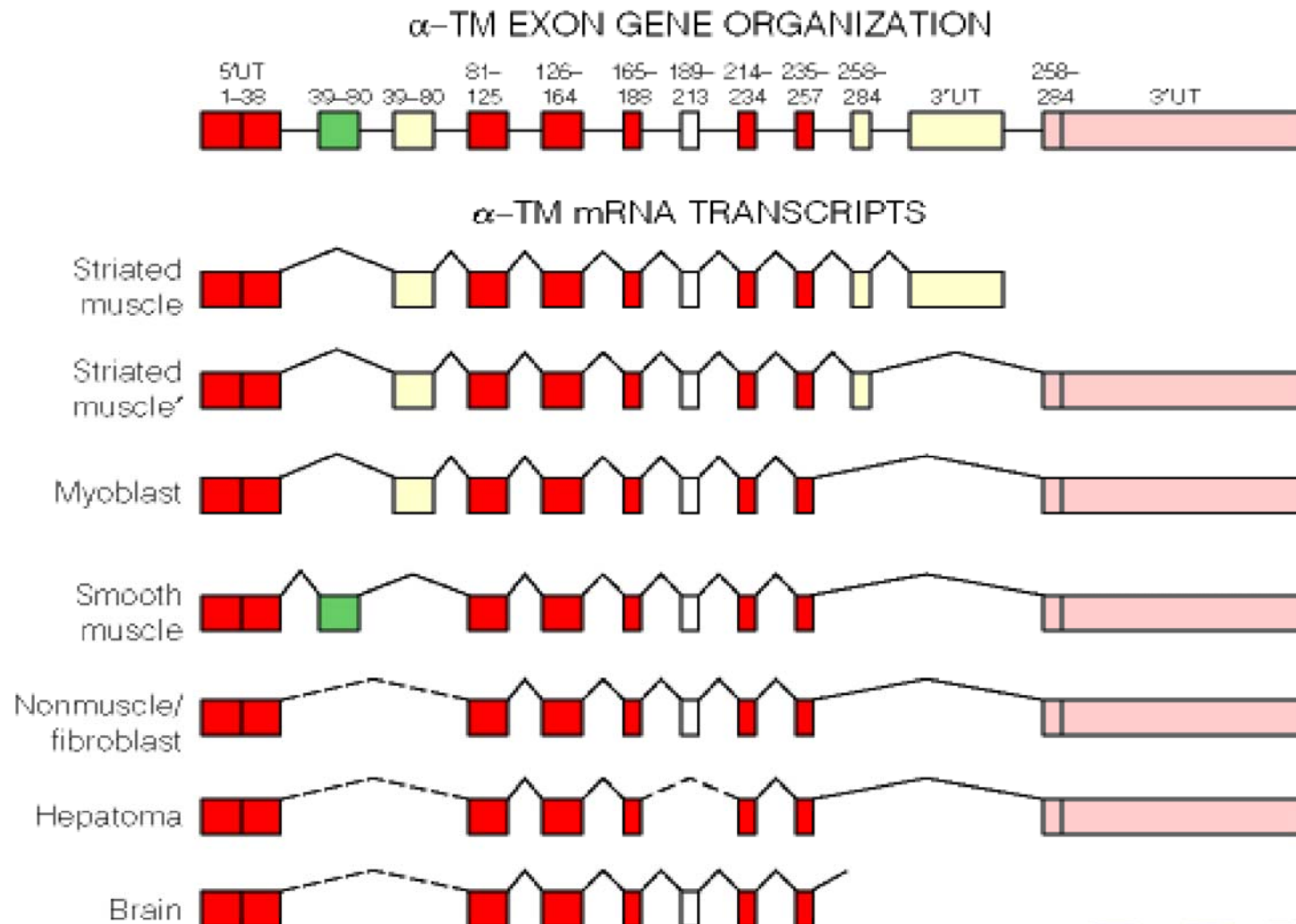
- What is dbEST?**: A section explaining that dbEST (Nature Genetics 4:332-3:1993) is a division of GenBank containing sequence data and other information on "single-pass" cDNA sequences, or "Expressed Sequence Tags", from a number of organisms. It also mentions a brief account of the history of human ESTs in GenBank is available (Trends Biochem. Sci. 20:295-6:1995) and suggests consulting the special "Genome Directory" issue of Nature (vol. 377, issue 6547S, 28 September 1995).
- Other ways to access dbEST**: A section with the text "Other ways to access dbEST".
- How to submit data**: A section with the text "How to [submit data to dbEST](#)".
- Information on the current release**: A section with the text "[Number of ESTs](#) - dbEST summary by organism".

The left sidebar contains navigation links for NCBI, SITE MAP, Human Genome Resources, UniGene, Entrez Gene, and NCI\_CGAP.

<http://www.ncbi.nlm.nih.gov/dbEST/index.html>

# Nucleotide Database

## Alternative splicing



# Alternative Splicing Database

[Home](#)[Introduction](#)[User's Guide](#)[Downloads](#)[Publications](#)[Pygr](#)[Forum](#)[Leelab](#)

## ASAP II

Select Organism & Database, and Type Input Query. Press SUBMIT to see output.	
Organism & Database	<input type="text" value="Human JAN06"/>
User Query	<input type="text"/>
	<input type="button" value="SUBMIT"/> <input type="button" value="RESET"/>

Supported query types:

- Gene Symbol: search by a gene symbol (e.g. BRCA1)
- UniGene Sequence Identifier: search by a UniGene sequence identifier (e.g. Hs#S20337840)
- UniGene Cluster Identifier: search by a UniGene cluster identifier (e.g. Hs.194143)
- Gene Title: search by a gene title (e.g. breast cancer 1)
- GeneBank Sequence Identifier: search by a GeneBank sequence identifier (e.g. BC046142)

Copyright (c) 2006 All Rights Reserved. Contact [Namshin Kim](#) and [Christopher Lee](#) for more information.



網際網路

<http://bioinformatics.ucla.edu/ASAP2/>

**SUN Lab**  
Molecular Genetics & Bioinformatics

# Nucleotide Database

## Single Nucleotide Polymorphism (SNP)

seq_1 (A)	ATGCGGC <b>G</b> ATTGCCATGGGTA
seq_2 (A)	ATGCGGC <b>G</b> ATTGCCATGGG <b>AA</b>
seq_3 (A)	ATGCGGC <b>G</b> ATTGCCATGGGTA
seq_1 (B)	ATGCGG <b>CA</b> ATTGCCATGGGTA
seq_2 (B)	ATGCGG <b>CA</b> ATTGCCATGGG <b>T</b>
seq_3 (B)	ATGCGG <b>CA</b> ATTGCCATGGGTA
Contig	ATGCGGC <b>G</b> ATTGCCATGGGTA

SNP ↑

↑ ↑

sequencing errors or paralogs

# SNP Database

The screenshot shows the dbSNP Home Page in a Windows Internet Explorer browser window. The address bar displays <http://www.ncbi.nlm.nih.gov/SNP/>. The page features the NCBI logo and the title "Single Nucleotide Polymorphism" with a molecular structure image. A navigation menu includes PubMed, Nucleotide, Protein, Genome, Structure, PopSet, Taxonomy, OMIM, Books, and SNP. A search bar is titled "Search for SNP on NCBI Reference Assembly" with a "Search Entrez" dropdown set to "SNP" and a "Go" button. A yellow "ANNOUNCEMENT" banner reads: "03/25/2010: RELEASE: NCBI dbSNP Build 131 for Human" and "dbSNP Build 131:human\_9606 is now available ([More...](#))". Below this is a "Search by IDs on All Assemblies" section with a note: "Note: **rs#** and **ss#** must be prefixed with 'rs' or 'ss', respectively (i.e. rs25, ss25)". It includes an "ID:" input field, a "Reference cluster ID(rs#)" dropdown, and "Search" and "Reset" buttons. A "Submission Information" section lists: "By Submitter", "New Submitted Batches", and "Method". A left sidebar contains links for "BUILD 131", "GENERAL", "HUMAN VARIATION", "SNP SUBMISSION", "DOCUMENTATION", "SEARCH", and "RELATED SITES". The Windows taskbar at the bottom shows the Start button, taskbar icons for Microsoft PowerPoint and PowerPoint 投影片放..., and the system tray with the time "下午 05:18".

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



# SNP Database



Welcome to the NCKU SNP Value-Added Database (SNP\_VAD)

## About SNP\_VAD

### Searching

- Text Search
- Blast Search

### Submission

- New PI
- New Submitter
- New Variants

- Update
- Withdraw

NCBI Submission

### Visitor Count

Month : 26  
Year : 1969  
Total : 4587

```
CGAACGAA  CGAA  ACGAA  AACGAAACGAA
ACGAACGAA  CGAAC  ACTCA  CACTCACTCAC
ACTCACTCAC  CTCAC  ACTCA  CACTCACTCAC
CGTTCTCGTT  TTCGT  GTTCC  CGTTCTCGTT
CGTTCTCGTT  TTCGT  GTTCC  CGTTCTCGTT
CGTTCTCGTT  TTCGT  GTTCC  CGTTCTCGTT
ACTCAC  CTCACTCACTCA  CACTC  CTCAC
GTGAGTGAG  TGAGTGAGTGAG  AGTGAGTGAGT
TGAGTGAGT  TGAGTGAGTGAG  AGTGAGTGAGT
ACGAACG  CGAACGAACGAA  AACGAACGAA
ACGAACG  CGAA  GAACGAA  AACGA
AGTGAG  GAGTG  TGAG  GAGTGAG  AGTGA
CGTTCTCGTT  TTCG  CGTTCC  CGTTCTCGTT
GTTCTCGTT  TTCG  CGTTCC  CGTTCTCGTT
CTCACTCAC  CTCAC  ACTCA  CACTC
```

Web Revised : August 1, 2007

<http://www.binfo.ncku.edu.tw/snp/>



# Protein Database

The screenshot shows a Windows Internet Explorer browser window displaying the ExPASy UniProt Knowledgebase website. The address bar shows the URL <http://tw.expasy.org/sprot/>. The browser's taskbar at the bottom shows several open applications: Microsoft PowerPoint, ExPASy - UniProt Knowledgebase, and PowerPoint 投影片放... (PowerPoint presentation). The website content includes the SIB logo (Swiss Institute of Bioinformatics) and the ExPASy logo. A search bar is visible with the text "UniProtKB" and "Go" and "Clear" buttons. The main heading is "ExpASY Proteomics Server". Below this, there are navigation links: "Databases", "Tools", "Services", "Mirrors", "About", and "Contact". The breadcrumb trail reads "You are here: ExpASY AU > Databases > Around UniProtKB". The main content area features the "swissprot" logo and the text "Swiss-Prot Protein knowledgebase" and "TrEMBL Computer-annotated supplement to Swiss-Prot". Below this, it states "The UniProt Knowledgebase consists of:" followed by a bulleted list:

- UniProtKB/Swiss-Prot; a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases [[More details / References / Linking to UniProtKB/Swiss-Prot / User manual / Recent changes / Disclaimer](#)].
- UniProtKB/TrEMBL; a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

Below the list, it says "These databases are developed by the Swiss-Prot groups at SIB and at EBI." There is a section for "UniProt Knowledgebase Release 2010\_06 consists of:" with two entries:

- UniProtKB/Swiss-Prot Release 2010\_06 of 18-May-10: 517100 entries ([More statistics](#))
- UniProtKB/TrEMBL Release 2010\_06 of 18-May-10: 10867798 entries ([More statistics](#))

To the right of these entries is a yellow box containing the text "> [Swiss-Prot headlines](#)  
[UniProt and Ensembl \(Read more...\)](#)". At the bottom of the main content area, there is a section titled "Access to the UniProt Knowledgebase".

<http://tw.expasy.org/sprot/>

# Protein Database

The screenshot shows a Windows Internet Explorer browser window displaying the ExPASy PROSITE website. The browser's address bar shows the URL <http://tw.expasy.org/prosite/>. The website header includes the SIB Swiss Institute of Bioinformatics logo and the ExPASy Proteomics Server logo. A search bar contains the text "PROSITE" and has "Go" and "Clear" buttons. The main content area features the PROSITE logo and the text "Database of protein domains, families and functional sites". Below this, there is a paragraph describing PROSITE and its associated tools. A section titled "PROSITE access" includes a search input field with the example text "e.g. PDOC00022, PS50089, SH3, zinc finger" and a "Search" button. To the right of the search field is a "Browse:" section with a list of options: "by documentation entry", "by ProRule description", "by taxonomic scope", and "by number of positive hit". Below the search section is a "PROSITE tools" section with a description of the "Scan a sequence against PROSITE patterns and profiles" tool and a list of tools: "ScanProsite", "PRATT", and "MyDomains - Image Creator". The browser's taskbar at the bottom shows several open applications, including Microsoft PowerPoint and ExPASy - PROSITE.

ExPASy - PROSITE - Windows Internet Explorer

<http://tw.expasy.org/prosite/>

Search PROSITE for  Go Clear

SIB Swiss Institute of Bioinformatics ExPASy Proteomics Server

Databases Tools Services Mirrors About Contact

You are here: ExPASy AU > Databases > PROSITE

Home ScanProsite ProRule Documents Downloads Links Funding

**prosite** Database of protein domains, families and functional sites

PROSITE consists of [documentation entries](#) describing protein domains, families and functional sites as well as associated [patterns](#) and [profiles](#) to identify them [[More details](#) / [References](#) / [Disclaimer](#) / [Commercial users](#)].  
PROSITE is complemented by [ProRule](#), a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [[More details](#)].

Release 20.64, of 18-May-2010 (1579 documentation entries, 1308 patterns, 888 profiles and 883 ProRule)

**PROSITE access**

e.g. PDOC00022, PS50089, SH3, zinc finger  Browse:

add wildcard \*\*

- [by documentation entry](#)
- [by ProRule description](#)
- [by taxonomic scope](#)
- [by number of positive hit](#)

**PROSITE tools**

**Scan a sequence against PROSITE patterns and profiles** - quick scan  
(Output includes graphical view and feature detection)

- [ScanProsite](#) - advanced scan
- [PRATT](#) - allows to interactively generate conserved patterns from a series of unaligned proteins.
- [MyDomains - Image Creator](#) **new** - allows to generate

<http://tw.expasy.org/prosite/>

# Protein Database

The screenshot shows a Windows Internet Explorer browser window displaying the ExPASy SWISS-2DPAGE website. The browser's address bar shows the URL <http://tw.expasy.org/ch2d/>. The website header includes the SIB Swiss Institute of Bioinformatics logo and the ExPASy Proteomics Server logo. A search bar is visible with the text "SWISS-2DPAGE" and "for" followed by a search button and a "Clear" button. The main content area features the "swiss2Dpage" logo and the text "SWISS-2DPAGE Two-dimensional polyacrylamide gel electrophoresis database". Below this, there is a paragraph describing the database's content and a release note for version 18.6, dated September 2006, which includes updates up to 31-January-2008. The release note lists the species included: human, mouse, *Arabidopsis thaliana*, *Dictyostellium discoideum*, *Escherichia coli*, *Saccharomyces cerevisiae*, and *Staphylococcus aureus* (N315). At the bottom of the page, there are two columns of links: "Access to SWISS-2DPAGE" and "SWISS-2DPAGE documents".

ExPASy Proteomics Server

Databases Tools Services Mirrors About Contact

You are here: ExPASy AU > Databases > SWISS-2DPAGE

**swiss2Dpage**

**SWISS-2DPAGE**  
Two-dimensional polyacrylamide gel electrophoresis database

**SWISS-2DPAGE** contains data on proteins identified on various 2-D PAGE and SDS-PAGE reference maps. You can locate these proteins on the 2-D PAGE maps or display the region of a 2-D PAGE map where one might expect to find a protein from UniProtKB/Swiss-Prot [[More details](#) / [References](#) / [Linking to SWISS-2DPAGE](#) / [Commercial users](#) / [Disclaimer](#)].

**Release 18.6, September 2006 and updates up to 31-January-2008 (contains 1265 entries in 36 reference maps from human, mouse, *Arabidopsis thaliana*, *Dictyostellium discoideum*, *Escherichia coli*, *Saccharomyces cerevisiae*, and *Staphylococcus aureus* (N315)).**

[[Search](#)][[Documents](#)][[Services](#)][[Software](#)][[Related servers](#)][[Other databases](#)][[Job openings](#)]

Access to SWISS-2DPAGE	SWISS-2DPAGE documents
<ul style="list-style-type: none"><li>• <a href="#">by description</a> (any word in the ID, DE, GN and KW lines)</li><li>• <a href="#">by accession number</a> (AC lines)</li><li>• <a href="#">by clicking on a spot</a>: select one of our 2-D PAGE or SDS-PAGE reference maps, click on a spot and then get the corresponding information from the SWISS-2DPAGE database.</li></ul>	<ul style="list-style-type: none"><li>• <a href="#">User manual</a></li><li>• <a href="#">Release notes</a> (September 26, 2006)</li><li>• <a href="#">FAQ (Frequently Asked Questions about SWISS-2DPAGE)</a></li><li>• <b>Protocols:</b><ul style="list-style-type: none"><li>◦ <a href="#">Technical information</a> about 2-D PAGE (IPG's, silver)</li></ul></li></ul>

<http://tw.expasy.org/ch2d/>

# Protein Database

The screenshot shows a Windows Internet Explorer browser window displaying the ExPASy ENZYME database. The browser's address bar shows the URL <http://tw.expasy.org/enzyme/>. The page header includes the SIB logo (Swiss Institute of Bioinformatics) and the ExPASy Proteomics Server logo. A search bar contains the word "ENZYME" and has "Go" and "Clear" buttons. Below the header, the breadcrumb trail reads "You are here: ExPASy AU > Databases > Enzyme". The main content area features the "swissprot" logo and the text "ENZYME Enzyme nomenclature database". A paragraph of text describes the database as a repository of information on enzyme nomenclature, based on IUBMB recommendations. It mentions that ENZYME now includes entries with preliminary EC numbers, such as EC 3.5.1.n3. Below this text is a section titled "Access to ENZYME" with several search options: "by EC number" (with four input fields and a "Search" button), "by enzyme class", "by description (official name) or alternative name(s)" (with one input field and a "Search" button), "by chemical compound", "by cofactor", and "by search in comments lines". At the bottom of the page, there is a "Documents" section. The browser's taskbar at the bottom shows the Windows Start button, several open applications including Microsoft PowerPoint, and the system clock showing 5:22 PM on May 19, 2010.

<http://tw.expasy.org/enzyme/>

# Protein Database

The screenshot shows the RCSB Protein Data Bank website in a Windows Internet Explorer browser window. The address bar displays the URL <http://www.pdb.org/pdb/home/home.do>. The browser's taskbar at the bottom shows several open applications, including Microsoft PowerPoint and the RCSB Protein Data Bank browser window. The website header features the RCSB PDB logo and the text "An Information Portal to Biological Macromolecular Structures". Below the header, there is a search bar with a "Search" button and a link to "Advanced Search". The main content area is titled "A Resource for Studying Biological Macromolecules" and contains several paragraphs of text describing the database's resources and tools. A "Molecule of the Month" section highlights Epidermal Growth Factor (EGF) with a 3D molecular model. The right sidebar includes a "Customize This Page" section with options for "New Features" and "Latest Structures Widget", and a "News" section with a "Statement on Retraction of PDB Entries" dated 2010-06-08. The browser's status bar at the bottom indicates the time as 下午 05:23.

<http://www.pdb.org/pdb/home/home.do>

# Other-type Databases

## GeneCards

The screenshot displays the GeneCards V3 website interface. The browser window title is "GeneCards V3 - Human Genes | Gene Database | Gene Search - Windows Internet Explorer". The address bar shows "http://www.genecards.org/". The page features a search bar with the text "Search the GeneCards human gene database" and a search button. Below the search bar, there are links for "Advanced Search" and "About V3 Search". The main content area is divided into sections: "GeneCards Suite" (listing GeneDecks, GeneALaCart, GeneLoc, GeneNote, GeneAnnot, and GeneTide), "Explore GeneCards" (with a description of the database and links for "GeneALaCart", "Hot genes", and "Disease genes"), and "View Sample Gene" (showing details for PSEN2, including aliases, pathways, SNPs, and other data). A "View Random Gene" section is also visible, showing details for AMBP. The right sidebar contains "News and Views" with links to "Version 3.02 features", "New search engine", and "What's New". The bottom of the page shows the Windows taskbar with various open applications and the system clock.

<http://bioinformatics.weizmann.ac.il/cards/>

# Other-type Databases

## BIOCARTA

**BIOCARTA** HOME | ABOUT | SUPPORT | CONTACT US | CAREERS | NEWS  
Charting Pathways of Life

FEATURES PATHWAYS CUSTOM SERVICES GENES PRODUCTS LOGIN

**PATHWAYS > All Pathways**

**Observe how genes interact** in dynamic graphical models. Our online maps depict molecular relationships from areas of active research. In an "open source" approach, this community-fed forum constantly integrates emerging proteomic information from the scientific community. It also catalogs and summarizes important resources providing information for over 120,000 genes from multiple species. Find both classical pathways as well as current suggestions for new pathways.

**ALL PATHWAYS**

**A**

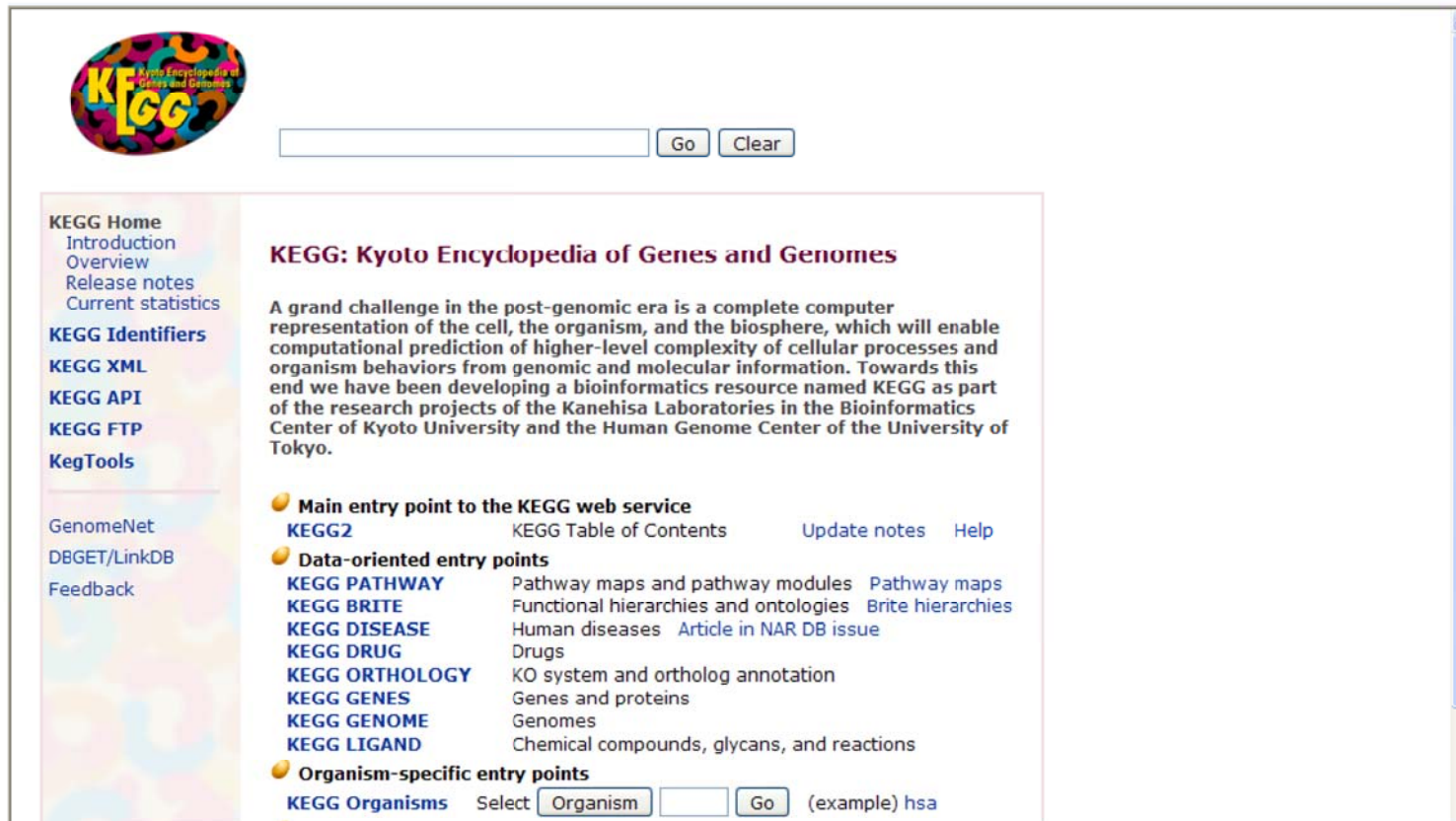
- ▶ Acetylation and Deacetylation of RelA in The Nucleus [H](#) [M](#)
- ▶ Actions of Nitric Oxide in the Heart [H](#)
- ▶ Activation of cAMP-dependent protein kinase, PKA [H](#) [M](#)
- ▶ Activation of Csk by cAMP-dependent Protein Kinase Inhibits Signaling through the T Cell Receptor [H](#) [M](#)
- ▶ Activation of PKC through G protein coupled receptor [H](#) [M](#)
- ▶ Activation of Src by Protein-tyrosine phosphatase alpha [H](#) [M](#)
- ▶ Acute Myocardial Infarction [H](#)
- ▶ Adhesion and Diapedesis of Granulocytes [H](#)
- ▶ Adhesion and Diapedesis of Lymphocytes [H](#)
- ▶ Adhesion Molecules on Lymphocyte [H](#) [M](#)
- ▶ ADP-Ribosylation Factor [H](#) [M](#)
- ▶ Agrin in Postsynaptic Differentiation [H](#) [M](#)
- ▶ Ahr Signal Transduction Pathway [H](#) [M](#)
- ▶ AKAP95 role in mitosis and chromosome dynamics [H](#) [M](#)
- ▶ AKT Signaling Pathway [H](#) [M](#)
- ▶ ALK in cardiac myocytes [H](#)
- ▶ Alpha-synuclein and Parkin-mediated proteolysis in Parkinson's disease [H](#) [M](#)
- ▶ Alternative Complement Pathway [H](#) [M](#)
- ▶ Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling [H](#) [M](#)
- ▶ Angiotensin-converting enzyme 2 regulates heart function [H](#) [M](#)
- ▶ Anthrax Toxin Mechanism of Action [H](#)
- ▶ Antigen Dependent B Cell Activation [H](#) [M](#)
- ▶ Antigen Processing and Presentation [H](#)
- ▶ Antisense Pathway [H](#) [M](#)

<http://www.biocarta.com/genes/allPathways.asp>



# Other-type Databases

## Kyoto Encyclopedia of Genes and Genomes (KEGG)



The screenshot shows the KEGG website homepage. At the top left is the KEGG logo, a colorful circle with the letters 'KEGG' inside. To its right is a search bar with 'Go' and 'Clear' buttons. Below the logo is a navigation menu with links for 'KEGG Home', 'Introduction', 'Overview', 'Release notes', 'Current statistics', 'KEGG Identifiers', 'KEGG XML', 'KEGG API', 'KEGG FTP', and 'KegTools'. Below the menu are links for 'GenomeNet', 'DBGET/LinkDB', and 'Feedback'. The main content area features the title 'KEGG: Kyoto Encyclopedia of Genes and Genomes' and a paragraph describing the database's mission. Below this are three sections: 'Main entry point to the KEGG web service' with links for 'KEGG2', 'Update notes', and 'Help'; 'Data-oriented entry points' with links for 'KEGG PATHWAY', 'KEGG BRITE', 'KEGG DISEASE', 'KEGG DRUG', 'KEGG ORTHOLOGY', 'KEGG GENES', 'KEGG GENOME', and 'KEGG LIGAND'; and 'Organism-specific entry points' with a 'KEGG Organisms' section containing a 'Select' dropdown, an 'Organism' input field, a 'Go' button, and the example '(example) hsa'.

**KEGG: Kyoto Encyclopedia of Genes and Genomes**

A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. Towards this end we have been developing a bioinformatics resource named KEGG as part of the research projects of the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo.

- **Main entry point to the KEGG web service**
  - KEGG2      KEGG Table of Contents      Update notes      Help
- **Data-oriented entry points**
  - KEGG PATHWAY      Pathway maps and pathway modules      Pathway maps
  - KEGG BRITE      Functional hierarchies and ontologies      Brite hierarchies
  - KEGG DISEASE      Human diseases      Article in NAR DB issue
  - KEGG DRUG      Drugs
  - KEGG ORTHOLOGY      KO system and ortholog annotation
  - KEGG GENES      Genes and proteins
  - KEGG GENOME      Genomes
  - KEGG LIGAND      Chemical compounds, glycans, and reactions
- **Organism-specific entry points**
  - KEGG Organisms      Select                  (example) hsa

<http://www.genome.jp/kegg/>

# Other-type Databases

## Signaling Pathway Database

SPAD - Signaling Pathway Database - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

連結 Windows Windows Media 自訂連結 免費的 Hotmail

網址(D) http://www.grt.kyushu-u.ac.jp/spad/ 移至

← 上一頁 · → · 搜尋 · 我的最愛 · 媒體 · Norton AntiVirus


### Signaling Pathway Database

The Signaling Pathway Database (SPAD) is an integrated database for genetic information and signal transduction systems.

There are multiple signal transduction pathways: cascade of information from plasma membrane to nucleus in response to an extracellular stimulus in living organisms. Extracellular signal molecule binds specific intracellular receptor, and initiates the signaling pathway. Now, there is a large amount of information about the signaling pathway which controls the gene expression and cellular proliferation. We have developed an integrated database SPAD to understand the overview of signaling transduction. SPAD is divided to four categories based on extracellular signal molecules (Growth factor, Cytokine, and Hormone) and stress, that initiate the intracellular signaling pathway. SPAD is compiled in order to describe information on interaction between protein and protein, protein and DNA as well as information on sequences of DNA and proteins.

There are two methods for retrieving this database. Please select one of the two items.

- [Extracellular Signal Molecules](#)

 This WWW service "SPAD" is still under development.

mail to: [sachiyo@](mailto:sachiyo@)

**Molecular Gene Technics**  
Hakozaki Higashi-ku,  
Fukuoka, 812-8581, Japan  
Graduate School of Genetic Resources Technology  
Kyushu University

2004年2月23日

開始 Microsoft PowerPoint ... 8 Internet Explorer 上午 05:29

<http://www.grt.kyushu-u.ac.jp/spad/>

# Other-type Databases

The screenshot shows a web browser window titled "CSHL Mammalian Promoter Database (CSHLmpd) - Windows Internet Explorer". The address bar contains "http://rulai.cshl.edu/CSHLmpd2/". The browser's menu bar includes "檔案(F)", "編輯(E)", "檢視(V)", "我的最愛(A)", "工具(T)", and "說明(H)". The address bar also features a "Live Search" field. The browser's toolbar includes icons for home, back, forward, print, and other functions. The main content area displays the CSHL logo and the title "Cold Spring Harbor Laboratory Mammalian Promoter Database (Version 2.0, May 2005)". Below the title is a navigation menu with tabs for "Databases", "Human", "Mouse", "Rat", and "TRED". The "Introduction" section is visible, containing text about the database's purpose and construction. A left sidebar contains links for "Home", "About CSHLmpd", "Resources", and "Related links". The browser's status bar at the bottom shows "網際網路" and "100%".

**CSHL Mammalian Promoter Database (CSHLmpd)**  
Cold Spring Harbor Laboratory  
Mammalian Promoter Database  
(Version 2.0, May 2005)

**Databases**    **Human**    **Mouse**    **Rat**    **TRED**

**Home**  
- [Current version](#)  
- [Old version](#)

**About CSHLmpd**  
- [Method](#)  
- [Statistics](#)  
- [References](#)  
- [Release Notes](#)  
- [Credits](#)

**Resources**  
- [DME](#)  
- [CREAD](#)  
- [TCAT](#)

**Related links**  
- [ZhangLab](#)  
- [CSHL](#)

**Introduction**

In the post-genome era, characterization of gene regulation networks has become an important part of genomic research. To succeed in such studies in any organism, a high-quality and comprehensive database of genes and their promoters, transcription factor binding sites, and other cis-regulatory elements is much desired if not a must.

Cold Spring Harbor Laboratory mammalian promoter database (CSHLmpd) used all known transcripts, integrating with predicted transcripts, to construct gene set of human, mouse and rat genomes. For promoter information, we collected known promoter information from multiple resources, together with predicted ones. These promoters were mapped to genome, and linked to related genes. We also compared promoters of orthologous gene groups to detect the sequence conservation in promoter regions.

We expect CSHLmpd to be helpful for research in gene regulation networks by providing guidance for experimental studies such as DNA microarray and chromatin IP. It will also facilitate the building of a foundation upon which we expand our insights into the structure of mouse genome through continued data collection, intelligent data analysis and integration.

<http://rulai.cshl.edu/CSHLmpd2/>

# Other-type Databases

Promoter Database - Windows Internet Explorer

http://rulai.csh.edu/cgi-bin/TRED/tred.cgi?process=home

Transcriptional Regulatory Element Database

HOME

Database Statistics

Browse Database

Retrieve Promoters

Search TF Target Genes

Retrieve TF Motifs

Gene Regulatory Networks

Search Orthologs

Sequence Analysis Tools

- Regular Expression Search
- Matrix Search
- Motif Finding (DWE)
- PromoterWise Alignment
- Palindrome search

Useful Links

- MEME
- FootPrinter
- Gibbs Sampler

Introduction

In order to understand gene regulation, accurate and comprehensive knowledge of transcriptional regulatory elements is essential. Transcriptional Regulatory Element Database (TRED) has been built in response to increasing needs of an integrated repository for both cis- and trans- regulatory elements in mammals, and the lack of such resources at present.

Genome-wide human, mouse and rat promoter annotation in TRED was realized by an automated pipeline to extract known promoters from databases such as Genbank, EPD and DBTSS, and employ promoter finding program FirstEF combined with mRNA/EST information and cross-species comparisons. We have also carried out hand curation to assess computational prediction and ensure data accuracy. A quality level is assigned to each promoter based on the reliability of the supporting evidence.

Curation has also been done for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. Binding motifs are mapped on promoters of the corresponding genes and binding quality levels are assigned based on definitiveness of the binding evidence. Curation is currently focusing on target genes of 36 cancer-related TF families.

Distinguishing features of TRED include:

- relatively complete genome wide promoter annotation for human, mouse and rat
- availability of transcription factor binding and regulation information
- data accuracy is ensured by curation, which continues to expand
- easy and flexible data retrieval
- availability of on-the-fly sequence analysis tools

TRED can provide good training datasets for further genome wide cis-regulatory element prediction, assist

[http://rulai.csh.edu/cgi-bin/TRED/tred.cgi?  
process=home](http://rulai.csh.edu/cgi-bin/TRED/tred.cgi?process=home)

# Other-type Databases

The screenshot shows a Windows Internet Explorer browser window displaying the ARE Database at KFSHRC website. The address bar shows the URL <http://brp.kfshrc.edu.sa/ARED/>. The page content includes the BRP Home logo, the title "The ARED Home Page", and the subtitle "AU-RICH ELEMENT-CONTAINING mRNA DATABASE". The page is organized into five columns of links:

- ARED 1.0**
  - [ARED 1.0 Article Abstract](#)
  - [Methodology](#)
  - [ARED 1.0 / Cluster Groups](#)
  - [Biological Diversity](#)
- ARED 2.0**
  - [ARED 2.0 Article](#)
  - [ARED 2.0 Database Search Engine](#)
  - Functional Genomics (Uses of ARED):**
    - [mRNA turnover studies](#)
    - [Other studies](#)
- ARED 3.0**
  - [What's new](#)
  - [ARED 3.0 Database Search Engine](#)
- ARED Integrated**
  - [What's new](#)
  - [ARED Integrated Search Engine](#)
- ARED Organism**
  - [What's new](#) (Updated! 5-Mar-2008)
  - [ARED Organism Search Engine](#)

At the bottom, there is contact information for Dr. Khalid S.A. Khabar ([khabar@kfshrc.edu.sa](mailto:khabar@kfshrc.edu.sa)) and a "Credits" link. The Windows taskbar at the bottom shows the Start button, taskbar icons for Microsoft PowerPoint and the ARE Database at KFSHRC, and the system tray with the time 05:32.

<http://brp.kfshrc.edu.sa/ARED/>

# Other-type Databases



## Tumor Associated Gene

- 
-  [Oncogenic Domain Analysis](#)
  -  [Text Search](#)
  -  [List by Chromosome](#)
  -  [Non-TAG gene search](#)
- 

The completion of human genome sequences allows one to rapidly identify and analyze genes of interest through the use of computational approach. The available annotations including physical characterization and functional domains of known tumor-related genes thus can be used to study the role of genes involved in carcinogenesis. The tumor-associated gene (TAG) database was designed to utilize information from well-characterized oncogenes and tumor suppressor genes to facilitate cancer research. All target genes were identified through text-mining approach from the PubMed database. A semi-automatic information retrieving engine was built to collect specific information of these target genes from various resources and store in the TAG database. At current stage, 476 TAGs including 164 oncogenes, 161 tumor suppressor genes, and 151 genes related to oncogenesis were collected. Information collected in TAG database can be browsed through user-friendly web interfaces that provide searching genes by chromosome or by keywords. The “consensus domain analysis” tool functions to identify conserved protein domains and GO terms among selected TAG genes. In addition, the “oncogenic domain analysis” can analyze oncogenic potential of any user-provided protein based on a weighed term frequency table calculated from the TAG proteins. This study was supported by grant from National research program for genomic medicine (NRPGM) and personnel from Bioinformatics Center of Center for Biotechnology and Biosciences in the National Cheng Kung University, Taiwan.

---

[http://www.binfo.ncku.edu.tw/TAG/  
GeneDoc.php](http://www.binfo.ncku.edu.tw/TAG/GeneDoc.php)

# Other-type Databases

The screenshot shows a Windows Internet Explorer browser window displaying the Oxford Journals website. The address bar shows the URL <http://www.oxfordjournals.org/nar/database/a/>. The page title is "Oxford Journals | Life Sciences | Nucleic Acids Research | Database Summary Paper Alpha List". The main content area features a large header for "Nucleic Acids Research" and a navigation menu with links for "ABOUT THIS JOURNAL", "CONTACT THIS JOURNAL", "SUBSCRIPTIONS", "CURRENT ISSUE", "ARCHIVE", and "SEARCH". Below the header, there is a breadcrumb trail: "Oxford Journals > Life Sciences > Nucleic Acids Research > Database Summary Paper Alpha List". The main heading is "2012 NAR Database Summary Paper Alphabetic List". To the right of this heading is a navigation box with links: "Compilation Paper", "Category List", "Alphabetical List", "Category/Paper List", and "Search Summary Papers". Below the heading is a row of letters: "1 2 3 4 5 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z". The first entry is "16S and 23S Ribosomal RNA Mutation Database" by Triman K.L., with links for "database" and "summary". The second entry is "2D-PAGE" by Pleissner, K.-P., Eifert, T., Buettner, S., Knipper, J., Schmelzer, P., Stein, R., Schmidt, F., Mattow, J., Zimny-Arndt, U., Schmid, M., Jungblut, P.R., with links for "database" and "summary". The third entry is "3D rRNA modification maps" with links for "database" and "summary". The fourth entry is "3D-Footprint" by Bruno Contreras-Moreira.

[http://www.oxfordjournals.org/nar/  
database/a/](http://www.oxfordjournals.org/nar/database/a/)

# Internet Resources

## Analytic tools

- **DNA/RNA**
  - Structure prediction
  - Translation
  - Open reading frame (ORF) identification
  - Pair-wise comparison
  - Multiple sequence comparison
  - Patterns and motifs
- **Protein**
  - Domain analysis
  - Structure
  - UTRs
  - .....



# DNA/RNA Structure Prediction

The Rensselaer bioinformatics web server

RNA & DNA Folding Applications

Home Applications

RNA & DNA Folding Applications

Download UNAFold software

Download mfold software

RNA Folding

DNA Folding

Nucleic Acid Quikfold

Zipfold Server

Tm Server

2-state hybridization server

Free Energy Determination

View Folding Results

Michael Zuker, professor of mathematical sciences, develops tools for predicting the secondary structure of RNA and DNA, mainly by using thermodynamic methods. Much of his work has been on RNA structure, which is important in understanding many biological processes, including translation regulation in messenger RNA, replication of single-stranded RNA viruses, and the function of structural RNAs and RNA/protein complexes. His algorithms have been widely used for drug design, and work on DNA folding has been very popular with the biotechnology community. Recent work in his laboratory includes the development of methods to predict folding hybridization and melting curves for two strands of RNA or DNA, and he is developing statistically based rules for RNA folding. His algorithms are available on this website, which is so popular that the server registers as many as 800,000 hits a month. His papers outlining his algorithms get cited almost every day of the year.

Nick Markham (Ph.D. 2006), a computer scientist who worked with professor Zuker, created the UNAFold software package that both replaces and extends mfold. It could be named mfold<sup>++</sup>. His DINAMelt web server, created in July 2005, is based on UNAFold. By the end of 2005, he had recreated the Quikfold, Zipfold, T<sub>m</sub> and 2-state hybridization servers and incorporated them into DINAMelt.

**Current applications:**

- Download UNAFold software
- Download mfold software
- RNA Folding
- DNA Folding
- "Bulk" servers: Fast results for many (pairs of) sequences
  1. Nucleic Acid Quikfold

Michael Zuker is a Professor of Mathematical Sciences at Rensselaer Polytechnic Institute, School of Science. Professor Zuker is interested in bioinformatics, specifically in algorithms for nucleic acid and protein sequence analysis. He is best known for his work on algorithms for predicting RNA and DNA secondary structure.

Michael Zuker  
Professor of Mathematical Sciences  
331 Amos Eaton Hall  
Rensselaer Polytechnic Institute  
110 8<sup>th</sup> Street  
Troy, NY 12180-3590

Rensselaer  
Tel: (518) 276-6002 (6345)  
Fax: (518) 275-4834  
E-mail: [mzk@rpi.edu](mailto:mzk@rpi.edu)

<http://frontend.bioinfo.rpi.edu/applications/mfold/>

# Translation Tool

The screenshot shows a Windows Internet Explorer browser window titled "ExPASy - Translate tool - Windows Internet Explorer". The address bar contains the URL <http://tw.expasy.org/tools/dna.html>. The browser's taskbar at the bottom shows several open applications: "開始" (Start), "Microsoft PowerPoint...", "ExPASy - Translate to...", and "PowerPoint 投影片放...". The system clock in the bottom right corner of the taskbar displays "下午 05:39".

The web page content includes the SIB Swiss Institute of Bioinformatics logo and the ExPASy Proteomics Server logo. A search bar is located at the top right with the text "Search ExPASy web site for" and "Go" and "Clear" buttons. A navigation menu contains links for "Databases", "Tools", "Services", "Mirrors", "About", and "Contact". A breadcrumb trail reads "You are here: ExPASy AU > Tools > DNA -> Protein > Translate".

The main heading is "Translate tool". Below it, a description states: "Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence." A text input area is provided with the instruction: "Please enter a DNA or RNA sequence in the box below (numbers and blanks are ignored)." Below the input area, the "Output format" is set to "Verbose ('Met', 'Stop', spaces between residues)". At the bottom of the tool, there are "Reset" and "TRANSLATE SEQUENCE" buttons.

<http://tw.expasy.org/tools/dna.html>

# Translation Tool

Transeq Nucleotide to Protein Sequence Conversion | EBI - Windows Internet Explorer

http://www.ebi.ac.uk/Tools/emboss/transeq/index.html

EMBL-EBI

EB-eye Search All Databases Enter Text Here Go Reset Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

Help Index  
General Help  
Formats  
Gaps  
Matrix  
References  
EMBOSS-Transeq Help  
Emboss Programmatic Access

EBI > Tools > Sequence Analysis > EMBOSS

### EMBOSS Transeq

Transeq translates nucleic acid sequences to the corresponding peptide sequence. It can translate in any of the 3 forward or three reverse sense frames, or in all three forward or reverse frames, or in all six frames.

Frame: 1 Table: Standard Code

Regions: START-END Trim: No Reverse: No Colour: No

Enter or Paste a nucleic acid Sequence in any format:  Help

Upload a file:  瀏覽...

If you plan to use these services during a course please [contact us](#).

Microsoft PowerPoint ... Transeq Nucleotide to ... 下午 05:40

<http://www.ebi.ac.uk/emboss/transeq/>

# Back-translation Tool

EMBOSS Backtranseq | Sequence Translation Sites | EBI - Windows Internet Explorer

http://www.ebi.ac.uk/Tools/st/emboss\_backtranseq/

EMBL-EBI

Enter Text Here Find Terms of Use Privacy Cookies

Databases Tools Research Training Industry About Us Help Site Index

EMBOSS Backtranseq

EMBOSS Backtranseq reads a protein sequence and writes the nucleic acid sequence it is most likely to have come from.

Use this tool

STEP 1 - Enter your input sequence

Enter or paste a set of Protein sequences in any supported format

Or, upload a file: [Browse...]

STEP 2 - Select Parameters

CODON USAGE TABLE

Homo sapiens

上午 05:44

[http://www.ebi.ac.uk/Tools/st/emboss\\_backtranseq/](http://www.ebi.ac.uk/Tools/st/emboss_backtranseq/)

# ORF Identification

The screenshot shows the NCBI ORF Finder web application. The browser window title is "ORF Finder - Windows Internet Explorer" and the address bar shows "http://www.ncbi.nlm.nih.gov/gorf/gorf.html". The page features a navigation bar with links to PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. A left sidebar contains links for NCBI, Tools for data mining, GenBank, and FTP site. The main content area includes a description of the ORF Finder tool, a form to "Enter GI or ACCESSION" with "OrfFind" and "Clear" buttons, a text area for "or sequence in FASTA format", "FROM:" and "TO:" input fields, and a "Genetic codes" dropdown menu set to "1 Standard". At the bottom, there are contact and credit information: "Comments and suggestions to: [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov)" and "Credits to: [Tatiana Tatusov](#) and [Roman Tatusov](#)". The Windows taskbar at the bottom shows the Start button, taskbar icons for Microsoft PowerPoint and ORF Finder, and a system tray with the time "下午 05:43".

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

# ORF Identification

The screenshot shows a Windows Internet Explorer browser window displaying the GENSCAN Web Server at MIT. The browser's address bar shows the URL <http://genes.mit.edu/GENSCAN.html>. The page content is on a dark blue background with white text. At the top, it reads "The GENSCAN Web Server at MIT" and "Identification of complete gene structures in genomic DNA". Below this, there is a green question mark icon followed by the text "For information about Genscan, click here". A "Server update, November, 2009" notice follows, stating that the webserver hardware was upgraded and some output errors were resolved. The main text describes the server's purpose: "This server provides access to the program Genscan for predicting the locations and exon-intron structures of genes in genomic sequences from a variety of organisms." It also mentions that the server can accept sequences up to 1 million base pairs (1 Mbp) in length. Below the text are several input fields and dropdown menus: "Organism:" with a dropdown menu set to "Vertebrate", "Suboptimal exon cutoff (optional):" with a dropdown menu set to "1.00", "Sequence name (optional):" with an empty text box, "Print options:" with a dropdown menu set to "Predicted peptides only", and "Upload your DNA sequence file (upper or lower case, spaces/numbers ignored):" with an empty text box and a "瀏覽..." button. The Windows taskbar at the bottom shows the Start button and several open applications, including Microsoft PowerPoint and the GENSCAN web browser. The system clock in the bottom right corner shows "下午 05:45".

<http://genes.mit.edu/GENSCAN.html>

# Sequence Comparisons

## Pair wise Sequence Alignment

- LALIGN (global & local)
- BLAST2 (local)
- DNADot (global)

## Multiple Sequence Alignments

- Multiple Sequence Alignments
- CHAOS/DIALIGN (Iterative alignment)
- Kalign

## List of sequence alignment tools

# Transcription Factor Prediction

**TESS : Transcription Element Search System**

Home AnGEL CRM Searches Site Searches Query Databases Other Stuff

Guide TESS CBIL PCBI UPenn

Welcome 140.116.141.66 *Need help? Check our FAQ page then please send questions and comments to [TessMaster@cbil.upenn.edu](mailto:TessMaster@cbil.upenn.edu).*

### Database Versions

TRANSFAC v6.0  
JASPAR 20060301  
IMD v1.1  
CBIL/GibbsMat v1.1

### Disclaimers

The TRANSFAC database is free for non-commercial use. For commercial use the TRANSFAC databases and programs have to be licensed. Please read the [DISCLAIMER!](#)

### News

**New Feature:** TESS now include **genome-wide CRM searches** and we've upgraded to **TRANSFAC v6.0** and we include **JASPAR** as well. 2007/02/28

**New Feature:** TESS has a new look and feel and we have added these new...

### Introduction

TESS is a web tool for predicting transcription factor binding sites in DNA sequences. It can identify binding sites using site or consensus strings and positional weight matrices from the TRANSFAC, JASPAR, IMD, and our CBIL-GibbsMat database. You can use TESS to search a **few of your own sequences** or for user-defined CRMs **genome-wide** near genes throughout genomes of interest.

### Search for CRMs Genome-wide

TESS now has the ability to search whole genomes for user defined CRMs. Try a search in the [AnGEL CRM Searches](#) section of the navigation bar.. You can search for combinations of consensus site sequences and/or PWMs from TRANSFAC or JASPAR.

Recall results of an earlier TESS-AnGEL CRM search:

TESS-AnGEL-

### Search DNA for Binding Sites

TESS also lets you search through your own sequence for TFBS. You can include your

Try a quick search with the default parameters

Job Title:

Email:

<http://www.cbil.upenn.edu/cgi-bin/tess/tess>





# The Binding Element Searching Tools


**NCKU Bioinformatics Center**

## THE **B**inding **E**lement **S**earching **T**ool

● Introduction ● Start ● My Result ● ReadMe

The **B**inding **E**lement **S**earching **T**ool is an online genome-wide searching tool.

Users can build up a customized hidden markov model (HMM) via HMMER program by upload or paste a set of element sequences containing specific transcription factor binding site (in FASTA format) and use this model to search the multi-species promoter region (human, mouse and rat) retrieved from the EnsMart system annotated by ensembl project. Results can be further analyzed by gene function and expression pattern via EnsMart interior linking to Gene Ontology and linkout to AFFY database.



© By [NCKU Bioinformatics Center](#), R.O.C. Government 2004. All Rights Reserved.

<http://thebest.binfo.ncku.edu.tw/thebest/>

# RNA Motif Prediction

RegRNA: A Regulatory RNA Motifs and Elements Finder - Windows Internet Explorer

http://regrna.mbc.nctu.edu.tw/

ISBLAB, Integrative Systems Biology Laboratory *Institute of Bioinformatics and Systems Biology, NCTU, Taiwan*

PI: Dr. Hsien-Da Huang [About Us](#) | [People](#) | [Research](#) | [Resource](#) | [Publications](#) | [mirrors](#)

Databases: miRNAMap dbPTM ProSplicer VITa RegRNA sRNAMap Tools: KinasePhos RNALogo PlantPAN miRStart RiboSW RNAMST

## RegRNA

A Regulatory RNA Motifs and Elements Finder

Release 1.0, Jan 2006

Home About Browse Prediction Statistics References Tutorial

» Welcome to RegRNA

**What's New**

- 01-06-2006
- Web site correction
- 02-09-2006
- AEdb integration
- 02-12-2006
- UTRSite updating
- 02-13-2006
- Web site correction
- 02-15-2006
- Riboswitches updating

**INFORMATION**

RegRNA is an integrated web server for identifying the homologs of **Regulatory RNA** motifs and elements against an input mRNA sequence. Both sequence homologs or structural homologs of regulatory RNA motifs can be identified. The regulatory RNA motifs supported in RegRNA are categorized into several classes:

- Motifs in mRNA 5'-UTR and 3'-UTR.
- Motifs involved in mRNA splicing.
- Motifs involved in transcriptional regulation.
- Other motifs in mRNA, such as riboswitches.
- Prediction of the splice sites, such as splicing donor/acceptor sites.
- RNA structural features, such as inverted repeat.
- miRNA target sites.

<http://regrna.mbc.nctu.edu.tw/>

# ExPASy Proteomics Server

The screenshot shows a Windows Internet Explorer browser window displaying the ExPASy Proteomics Server website. The browser's address bar shows the URL <http://tw.expasy.org/tools/>. The website header includes the SIB Swiss Institute of Bioinformatics logo and the ExPASy Proteomics Server title. A search bar is present with the text "ExPASy web site" and "for" followed by a search input field and "Go" and "Clear" buttons. The main content area is titled "ExPASy Proteomics tools" and contains a list of tools categorized under "Protein identification and characterization".

**ExPASy Proteomics tools**

The tools marked by are local to the ExPASy server. The remaining tools are developed and hosted on other servers.

[Protein identification and characterization] [Other proteomics tools] [DNA -> Protein] [Similarity searches] [Pattern and profile searches] [Post-translational modification prediction] [Topology prediction] [Primary structure analysis] [Secondary structure prediction] [Tertiary structure] [Sequence alignment] [Phylogenetic analysis] [Biological text analysis]

**Protein identification and characterization**

**Identification and characterization with peptide mass fingerprinting data**

- **Aldente** - Identify proteins with peptide mass fingerprinting data. A new, fast and powerful tool that takes advantage of Hough transformation for spectra recalibration and outlier exclusion. [Download the stand-alone version](#)
- **FindMod** - Predict potential protein post-translational modifications and potential single amino acid substitutions in peptides. Experimentally measured peptide masses are compared with the theoretical peptides calculated from a specified Swiss-Prot entry or from a user-entered sequence, and mass differences are used to better characterize the protein of interest.
- **FindPept** - Identify peptides that result from unspecific cleavage of proteins from their experimental masses, taking into account artefactual chemical modifications, post-translational modifications (PTM) and protease autolytic cleavage
- **GlycoMod** - Predict possible oligosaccharide structures that occur on proteins from their experimentally determined masses (can be used for free or derivatized oligosaccharides and for glycopeptides)

- **Mascot** - Peptide mass fingerprint from Matrix Science Ltd., London
- **PepMAPPER** - Peptide mass fingerprinting tool from UMIST, UK
- **ProFound** - Search known protein sequences with peptide mass information from Rockefeller and NY Universities [or from [Genomic Solutions](#)]
- **ProteinProspector** - UCSF tools for peptide masses data (MS-Fit, MS-Pattern, MS-Digest, etc.)

**Identification and characterization with MS/MS data**

<http://tw.expasy.org/tools/>

# Protein Pattern and Profile Searches

- **InterproScan**  
<http://www.ebi.ac.uk/InterProScan/>
- **SMART**  
<http://smart.embl-heidelberg.de/>
- **MOTIF**  
<http://motif.genome.jp/>

# Post-translational Modification Prediction

- **SignalP**  
<http://www.cbs.dtu.dk/services/SignalP/>
- **SecretomeP**  
<http://www.cbs.dtu.dk/services/SecretomeP/>
- **NetPhosK**  
<http://www.cbs.dtu.dk/services/NetPhosK/>

# Subcellular location Prediction



**PSORT**

<http://www.psort.org/>



**CELLO**

<http://cello.life.nctu.edu.tw/>



**TargetP**

<http://www.cbs.dtu.dk/services/TargetP/>

# Internet Resources

## Important bioinformatics centers

- National Center for Biotechnology Information  
(NCBI; <http://www.ncbi.nlm.nih.gov/>)
- European Bioinformatics Institute  
(EBI ; <http://www.ebi.ac.uk/>)
- Taiwan Bioinformatics Institute  
(TBI; <http://www.tbi.org.tw/>)
- The Center for Biological Sequence Analysis  
(CBS; <http://www.cbs.dtu.dk/services/>)



# The Center for Biological Sequence Analysis



CENTER FOR BIOLOGICAL SEQUENCE ANALYSIS ■ TECHNICAL UNIVERSITY OF DENMARK DTU



powered by sgi



<http://www.cbs.dtu.dk/services/>

**SUN Lab**  
Molecular Genetics & Bioinformatics

# Taiwan Bioinformatics Institute

The screenshot shows a Windows Internet Explorer browser window displaying the website <http://www.tbi.org.tw/>. The browser's address bar shows the URL, and the page title is "TBI Core Facility: - Windows Internet Explorer". The website's header features the TBI logo (a stylized DNA double helix) and the text "台灣生物資訊核心設施 Taiwan Bioinformatics Institute Core Facility" next to a molecular model. Below the header is a navigation menu with links: "About Us", "Customized Service", "Tools", "Databases", "Education", "Web Resources", "Newsletter", and "Help". The main content area is titled "What's New" and contains three news items:

- 「miRNA分析研習會」及「結構生物資訊學研習會」，將在6/26及6/27於「中國醫藥大學互助大樓7樓第一電腦教室」舉辦，歡迎報名參加！課程內容請參考[網頁資訊](#)！（公佈日期 2012/06/11）
- 「TBI Core Facility 電子報」第2期出刊，本期內容為 TBI - microRNA分析工具[歡迎點閱](#)！（公佈日期 2011/12/30）
- TBI核心設施將於100年11月9日（星期三，整天）- 11月10日（星期四，上午）於清華大學生命科學二館B1華生講堂（演講廳）舉辦「2011 Systems Biology and Bioinformatics Symposium」，本次研討會主題包括：系統生物學技術最新發展、高

On the right side of the page, there are two news sections: "NHRI News" and "NTHU News", each with a thumbnail image of a news article. The browser's taskbar at the bottom shows the Windows Start button, several application icons (including Photoshop and Skype), and the system tray with the date and time "下午 03:24".

<http://www.tbi.org.tw/>

**SUN Lab**  
Molecular Genetics & Bioinformatics