Introduction to bio-molecular databases

Information Sources in Biotechnology

Japan
What is EMBL?

- Non-profit organization
- Part of the European Molecular Biology Laboratory
- Based on the Wellcome Trust Genome Campus near Cambridge, UK
Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, The Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom

Associate member state: Australia
EMBL is a basic research institute funded by public research monies from 20 member states.

- 1400 staff
- over 60 nationalities
Website

EBI Search
Search all databases and literature in one go

http://www.ebi.ac.uk/
Website

Databases cover biology, chemistry and literature

http://www.ebi.ac.uk/
Databases: molecules to systems

Genomes
- Ensembl

Gene expression
- Gene Expression Atlas

Protein families, motifs and domains
- InterPro

Nucleotide sequence
- ENA

Proteomes
- UniProt, PRIDE, IntEnz

Protein structure
- PDBe

Protein interactions
- IntAct

Pathways
- Reactome

Chemical entities
- ChEBI, ChEMBL

Systems
- BioModels
Website

EMBL-EBI European Bioinformatics Institute

Tools for searching and analysis

http://www.ebi.ac.uk/
Gene prediction

Primer design

Microarray analysis

Pattern finding

Nucleotide translation

Protein functional analysis

Chemical substructure searching

Sequence similarity searching

Sequence alignment

Sequence analysis

Structural analysis
Website

http://www.ebi.ac.uk/
Website

http://www.ebi.ac.uk/

Industry support for SMEs and larger companies
Simple EBI Search
EBI Search...by gene name

Search for **src** gene

http://www.ebi.ac.uk/
EBI Search...by gene name

Lists species with information summaries

Lists relevant entries in all EBI resources

query

query

NOTE:
- Gene & Protein (includes expression, structures, literature...) grouped by orthology (with synonyms)
- EMBL-EBI European Bioinformatics Institute
EBI Search...by gene name

Tabs organise data by:
- gene
- expression
- protein
- structure
- literature

Easy to change between species
Information from **Ensembl**:

- Gene sequence
- Location
- Sequence variations
- Orthologues...

Gene structure

*forward and reverse strand*
Ensembl

Chromosomes

Gene families

Pick a genome

Genomic alignments

Gene trees

Synteny

Genes

Variations

User Upload

Variation Effect Predictor
Ensembl Genomes

Genome portals for the five kingdoms of life

Interface uses Ensembl technology

Variation data for plant, metazoan and fungal species

Compare whole bacterial chromosomes

Pan-taxonomic comparative analysis
Expression studies from Gene Expression Atlas, view by:

- Disease state
- Cell type
- Compound treatment...
ArrayExpress

ArrayExpress Archive: browse experiments
Search by keyword
Spreadsheets describing the sample properties
Gene Expression Atlas

Atlas: browse changes in gene expression

Search by gene or biological condition

Gene page

Experiment page
EBI Search...by gene name

Information from **UniProt**:
- Function
- Gene Ontology
- Isoforms
- Sequence...

**InterPro** domain architecture

**IntAct** protein interaction data
UniProtKB

Data sources

- GO
- PRIDE
- InterPro
- IntAct
- IntEnz
- HAMAP
- RESID

Some data sources for annotation

- Functional info
- Protein identification data
- Protein families and domains
- Molecular interactions
- Enzymes
- Microbial protein families
- Post-translational modifications

Protein classification

- InterPro classification
- Signal prediction
- Transmembrane prediction
- Other predictions

UniProt

- Manual curation
- Literature-based annotation
- Sequence analysis
- Automated annotation
Reactome

Select a pathway

Compare pathways between species

Interaction put on pathway diagram

View reactions and events in detail

View expression values overlaid on a pathway
### View family classifications

Proteins matched - Mitogen-activated protein kinase kinase kinase, 12/13 (246). The following proteins are predicted to be part of this family:

<table>
<thead>
<tr>
<th>Protein</th>
<th>Protein name</th>
<th>Species</th>
<th>Family</th>
<th>Domain organisation</th>
<th>Length Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>A71170</td>
<td>Mitogen-activated protein kinase kinase kinase 13-8</td>
<td>Xenopus laevis (african clawed frog)</td>
<td>Nitogen-activated protein kinase kinase kinase, 12/13</td>
<td>961</td>
<td>No</td>
</tr>
<tr>
<td>A72172</td>
<td>Mitogen-activated protein kinase kinase kinase 10-A</td>
<td>Xenopus laevis (african clawed frog)</td>
<td>Nitogen-activated protein kinase kinase kinase, 12/13</td>
<td>961</td>
<td>No</td>
</tr>
<tr>
<td>A71034</td>
<td>Mitogen-activated protein kinase kinase kinase 13</td>
<td>Bos taurus (Bovine)</td>
<td>Nitogen-activated protein kinase kinase kinase, 12/13</td>
<td>966</td>
<td>No</td>
</tr>
</tbody>
</table>

### Compare methods of protein signature prediction

#### GO classification

**GO Term Prediction**

- Biological Process:
  - GO:0007165 signal transduction
  - GO:0046647 filopodium assembly
- Molecular Function:
  - GO:0017124 SH3 domain binding
  - GO:0005515 protein binding
  - GO:0008093 cytoskeletal adaptor activity
- Cellular Component:
  - GO:0005622 intracellular

#### View domain architectures of proteins

**Domain organisation** - Zinc finger, RING-type

This domain is found in proteins with the following domain organisation:

<table>
<thead>
<tr>
<th>Domain organisation</th>
<th>Number of proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7343</td>
</tr>
<tr>
<td></td>
<td>4838</td>
</tr>
<tr>
<td></td>
<td>1545</td>
</tr>
<tr>
<td></td>
<td>546</td>
</tr>
<tr>
<td></td>
<td>441</td>
</tr>
<tr>
<td></td>
<td>355</td>
</tr>
<tr>
<td></td>
<td>304</td>
</tr>
</tbody>
</table>
EBI Search...by gene name

Information from PDBe:
- Chain information
- Structural domains
- Citations...

View additional structures
View structure
Ligand data: PDBeChem

Browse PDBe data: PDBeXplore

Sequence mapping SIFTS

PDBePrints: at-a-glance information

Biological assemblies and surface comparison: PDBePisa

Sites and motifs: PDBeMotif

Fold comparison: PDBeFold

Summary data for all PDBe entries

the European hub of the worldwide PBD resource for biological structures
EBI Search...by gene name

Can print full summary of any page

Reviews

Keyword in title

Free full text

Patent

Curator-selected
Links paper to EBI databases

Text-mining tools

List of publications

Detailed information, including abstract, citations, references, and chemicals
EBI Search...by patent number

Search for patent WO0146262

http://www.ebi.ac.uk/
EBI Search...by patent number

- E.coli and mouse orthologs
- Bioinformatics resources
- Literature for WO0146262
- Sequence data for WO0146262
  - Includes list of additional annotation
  - Includes link to full paper
EBI provides free access to >100 databases

EBI provides specialised data analysis tools

EBI provides multiple search options

EBI provides annotation-enriched search results

http://www.ebi.ac.uk/
Help

Contacts:

http://www.ebi.ac.uk/support/