Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

UniProt Knowledgebase: a hub of integrated data

Michele Magrane
magrane@ebi.ac.uk

EMBL-EBI

EBI is an Outstation of the European Molecular Biology Laboratory.

UniProt Knowledgebase (UniProtKB)

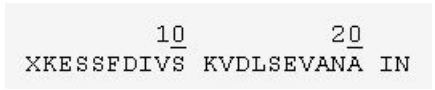
Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

- Produced by the UniProt Consortium

- Provides comprehensive overview of protein sequence and function
- Captures, interprets and incorporates data from range of sources

Sequence data

Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010



Submissions

Literature scanning

Annotation

Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

Manual curation

Literature

Sequence analysis

Automatic annotation

UniRule

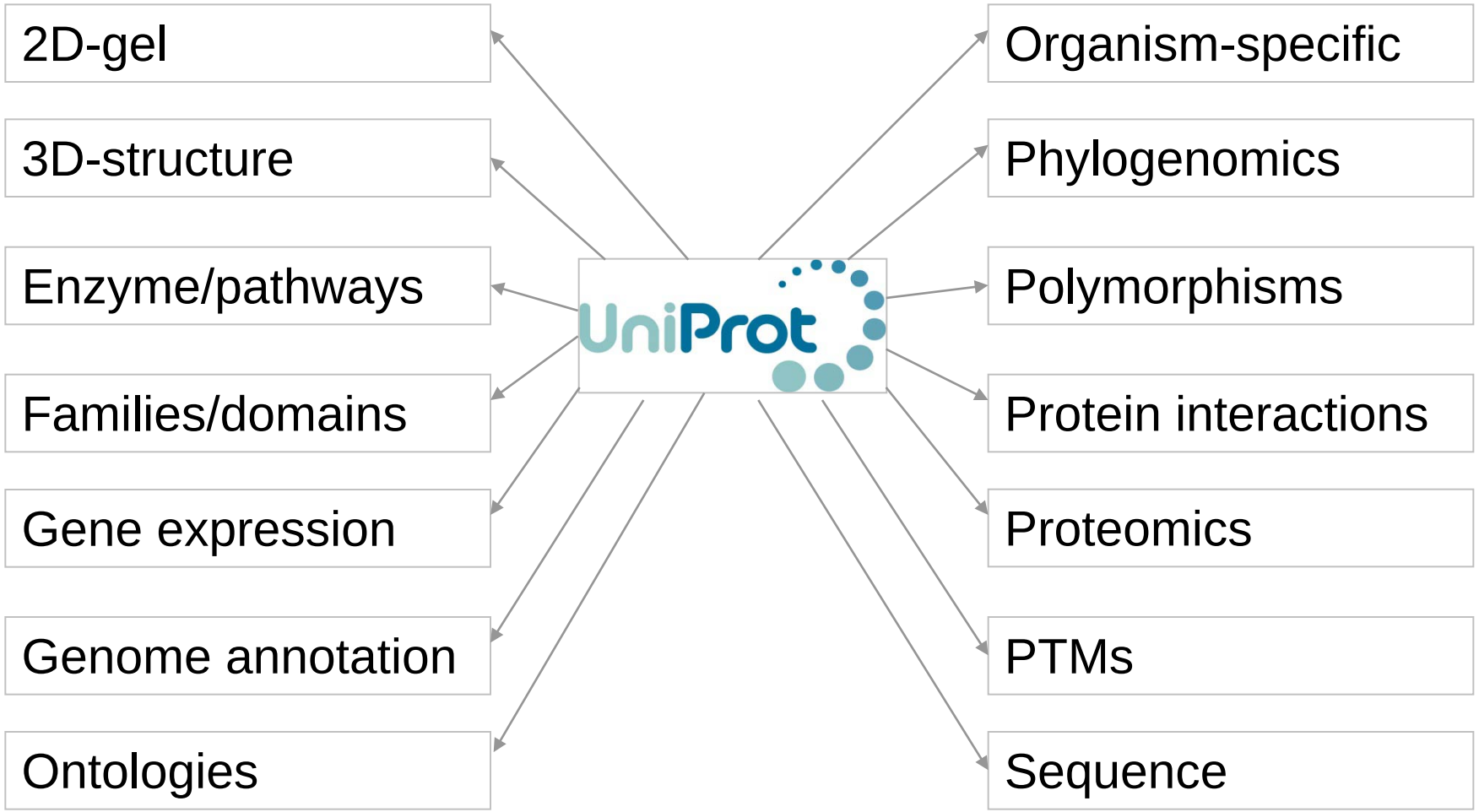
SAAS

UniProt

swissprot

TrEMBL
Translated EMBL

Cross-references



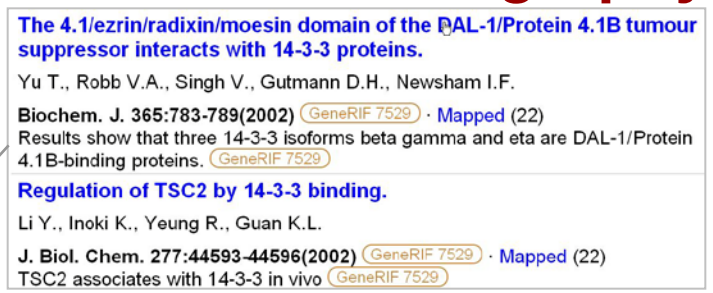
Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

Data import

Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

Gene names

Additional bibliography

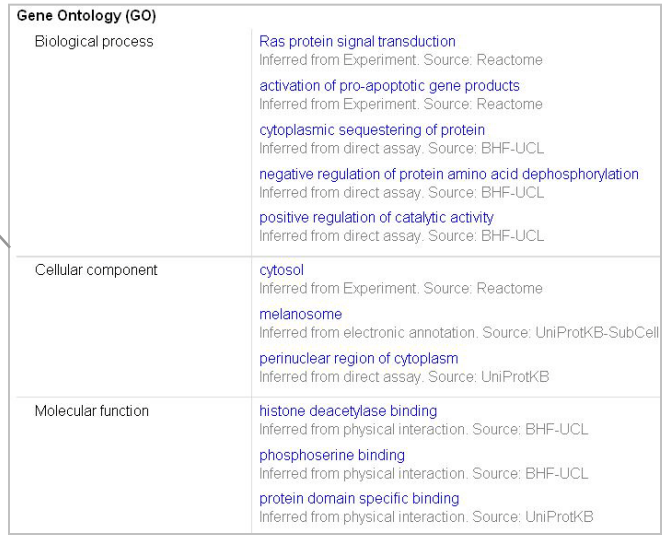


Interactions

Binary interactions

With	Entry	#Exp.	IntAct	Notes
itself		2	EBI-930964,EBI-930964	
ATXN2	Q99700	1	EBI-930964,EBI-697691	
BAT2	P48634	1	EBI-930964,EBI-347545	
C1orf94	Q6P1W5	1	EBI-930964,EBI-946029	
CCNK	O75909	1	EBI-930964,EBI-739806	
CFL1	P23528	1	EBI-930964,EBI-352733	
COIL	P38432	5	EBI-930964,EBI-945751	
CPSF7	Q8N684	1	EBI-930964,EBI-746909	
CRK	P46108	1	EBI-930964,EBI-886	
DAZAP2	Q15038	1	EBI-930964,EBI-724310	
DERP6	Q8TE02	1	EBI-930964,EBI-946189	

GO terms



DAS

P51587 (BRCA2_HUMAN) Reviewed, UniProtKB/Swiss-Prot

Last modified August 10, 2010. Version 122. History...

Contribute
Send feedback
Read comments (0) or add your own

Clusters with 100%, 90%, 50% identity | Documents (6) | Third-party data

text xml rdf/xml gff fasta

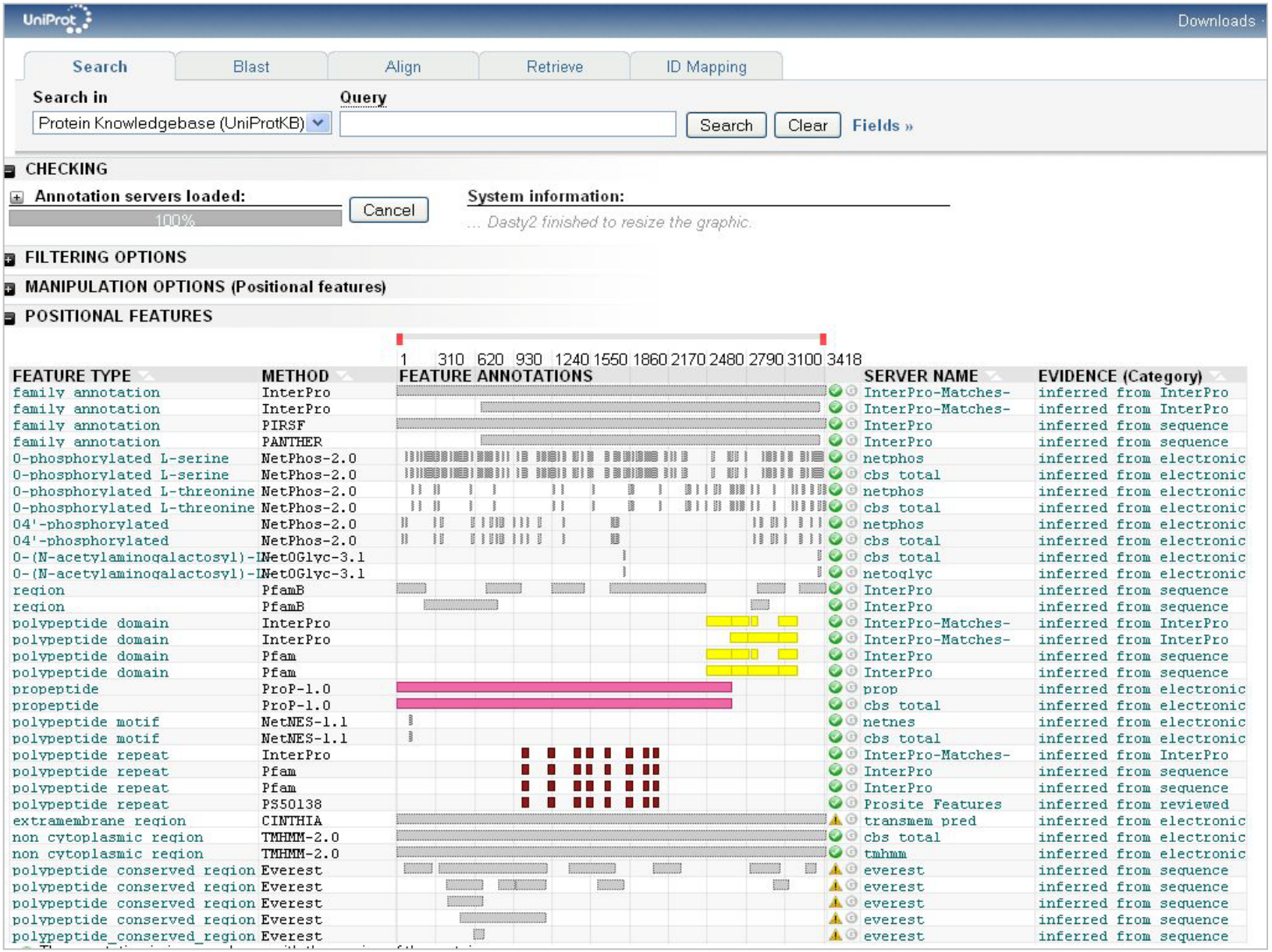
Customize display · Names · Attributes · General annotation · Ontologies · Interactions · Sequence annotation · Sequences · References · Web links · Cross-refs · Entry info · Documents

Names and origin

Protein names	*Recommended name:* **Breast cancer type 2 susceptibility protein** *Alternative name(s):* Fanconi anemia group D1 protein
Gene names	Name: **BRCA2** Synonyms:FACD, FANCD1
Organism	**Homo sapiens (Human)** [Complete proteome]
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo

Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010



Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

UniProt DAS server
http://www.ebi.ac.uk/uniprot-das/

The UniProt DAS Reference Server: Version 2.0

Welcome to the UniProt DAS Server that has been developed as a deliverable of the BioSapiens Network Project.

The UniProt DAS reference server is based upon the ***MyDas*** Java DAS Server Framework.

This new version launched on 05 of September of 2008 has several new features added to it. The UniProt data source is now capable of displaying keywords (KW), comments (CC). Ontologies have been added to the annotations so that these can be universaly identified.

A new data source has been added. The UniProt Spearmint allows the user to gain access to automatic annotations. These are predictions made on UniProt entries which are not accessible in the UniProt dsn. **The legacy aristotle data source has been discontinued and will be shutdown on at the end of this year. To access the entries previously serviced by aristotle, use one of the following: UniProt, UniParc and IPI.**

For an increase in performance we have removed access to UniProt data source via the remotingAPI UniProt Java Remoting API (UniProtJAPI). UniProtDAS now access the data source directly.

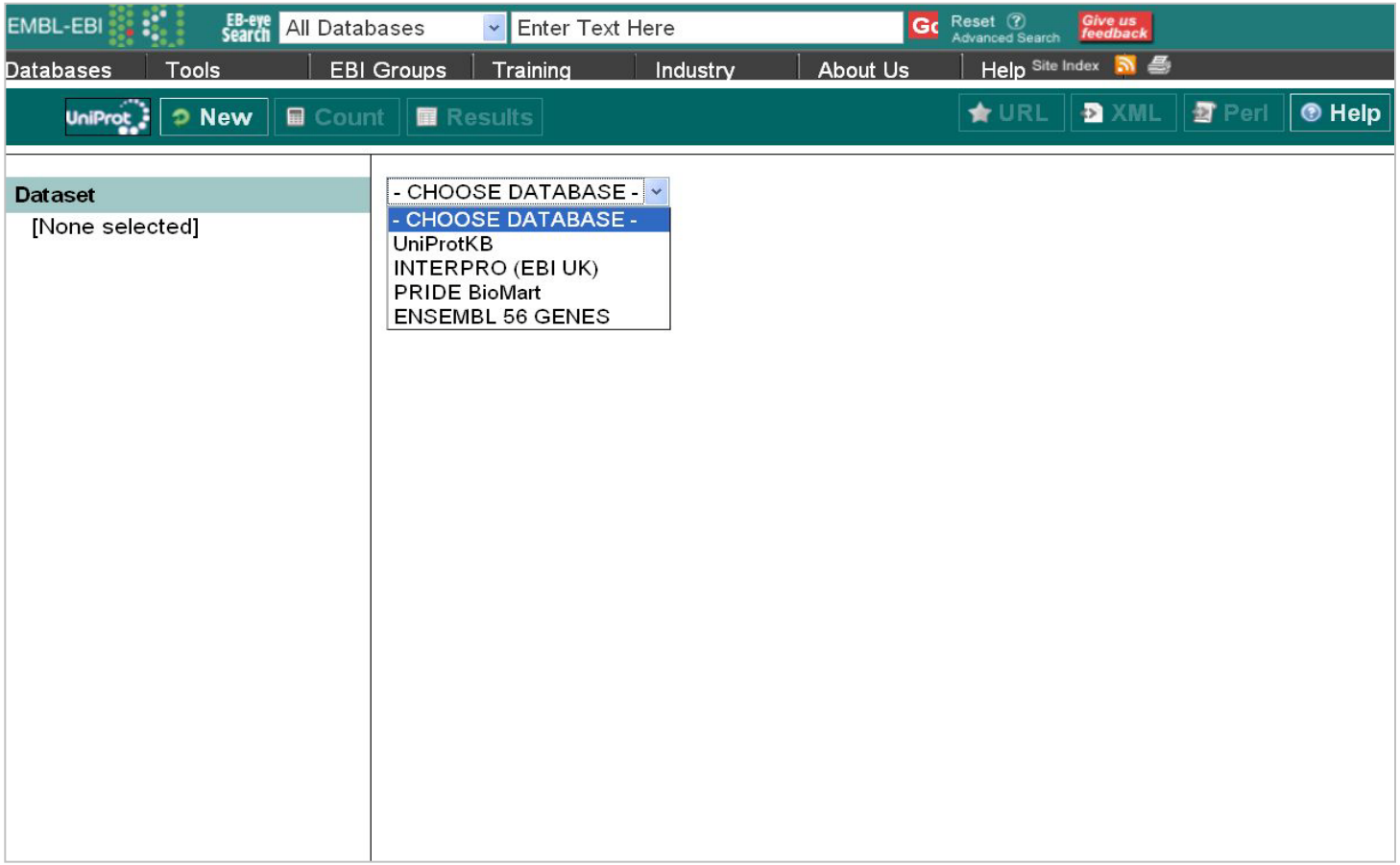
For details of the available data sources, please see the table below.

Data Sources

Short Name	Full Name	Version	Mapmaster	Description
ipi-tryptic	International Protein Index Predicted Tryptic Peptides	2.0	http://www.ebi.ac.uk/das-srv/uniprot/das/ipi-tryptic	Theoretical tryptic peptides based upon a complete and successful digest of the requested proteins with trypsin.
uniparc	UniProt Archive (UniParc)	2.0	http://www.ebi.ac.uk/das-srv/uniprot/das/uniparc	UniProt Archive (UniParc) is part of UniProt project. It is a non-redundant archive of protein sequences extracted from public databases UniProtKB/Swiss-Prot, UniProtKB/TrEMBL, PIR-PSD, EMBL, EMBL WGS, Ensembl, IPI, PDB, PIR-PSD, RefSeq, FlyBase, WormBase, H-Invitational Database, TROME database, European Patent Office proteins, United States Patent and Trademark Office proteins (USPTO) and Japan Patent Office proteins.
uniprot_summary	UniProt Summaries	1.0	http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot_summary	UniProt (Universal Protein Resource) is the world's most comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.
ipi	International Protein Index	2.0	http://www.ebi.ac.uk/das-srv/uniprot/das/ipi	IPI provides a top level guide to the main databases that describe the proteomes of higher eukaryotic organisms. IPI: 1. effectively maintains a database of cross references between the primary data sources 2. provides minimally redundant yet maximally complete sets of proteins for featured species (one sequence per transcript) 3. maintains stable identifiers (with incremental versioning) to allow the tracking of sequences in IPI between IPI releases. IPI is updated monthly in accordance with the latest data released by the primary data sources.
uniprot-spearmint	UniProt Spearmint	2.0	http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot-spearmint	Indicates the spearmint predictions for the uniprot entries
uniprot-goa	GO Annotation of UniProt	2.0	http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot-goa	The GOA project aims to provide high-quality Gene Ontology (GO) annotations to proteins in the UniProt Knowledgebase (UniProtKB) and International Protein Index (IPI) and is a central dataset for other major multi-species databases; such as Ensembl and NCBI.
uniprot-tryptic	UniProt Predicted Tryptic Peptides	2.0	http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot-tryptic	Theoretical tryptic peptides based upon a complete and successful digest of the requested proteins with trypsin.
uniprot	UniProt	2.0	http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot	UniProt (Universal Protein Resource) is the world's most comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.

UniProt BioMart

http://www.ebi.ac.uk/uniprot/biomart/martview



Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

UniProtKB data

- Released every 4 weeks and is freely available
  - http://www.uniprot.org/downloads
- Provided in a range of formats
  - text, XML, XML/RDF, FASTA, GFF, tab-delimited
- Web site
  - Supports simple and complex queries
- Data can be retrieved programmatically using REST
  - http://www.uniprot.org/faq/28
- UniProtJAPI
  - http://www.ebi.ac.uk/uniprot/remotingAPI/

Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

Future plans

- Curation
  - Provision of high-quality manual curation
  - Expand coverage of automatic annotation
- Cross-references
  - Maintain and expand linked resources
- Data import
  - Sequence data from RefSeq
  - Variant data from Ensembl
  - wwPDB data
  - Proteomics data

Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

Summary

- UniProt approach to data integration ensures that information is captured in the most appropriate resource for subsequent integration with other databases

- Maximum curation efficiency by preventing duplication of efforts across multiple resources

Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010

Acknowledgements

- UniProt curators at EBI, SIB and PIR
- UniProt programmers at EBI, SIB and PIR
- UniProt Consortium is led by Rolf Apweiler, Ioannis Xenarios, Cathy Wu
- Funders – NIH, EU, Swiss Government, NSF

Nature Precedings : doi:10.1038/npre.2010.5092.1 : Posted 22 Oct 2010