



UniProt Knowledgebase: a hub of integrated data

Michele Magrane
magrane@ebi.ac.uk



UniProt Knowledgebase (UniProtKB)

- Produced by the UniProt Consortium



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

Sequence data



10 20
XKESFFDIVS KVDLSEVANA IN

Submissions

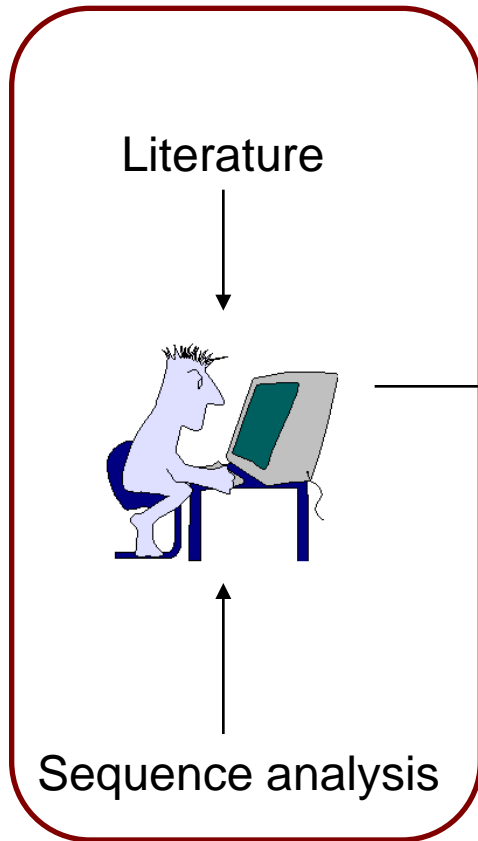


Literature scanning

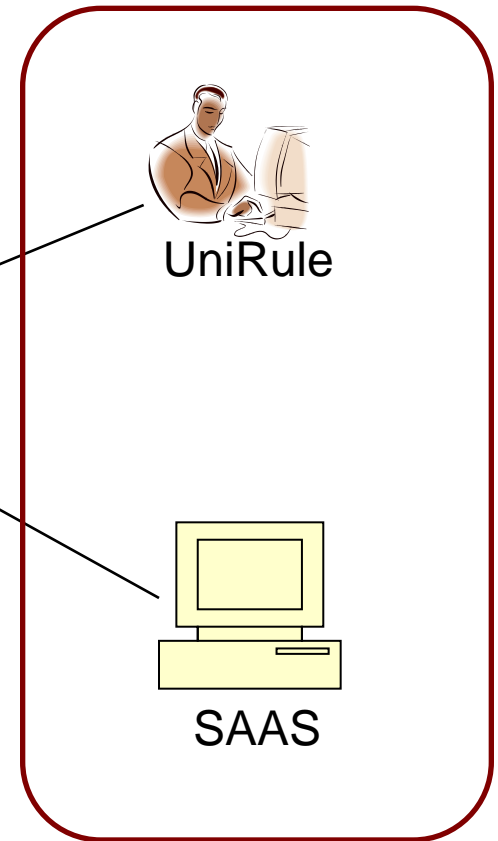


Annotation

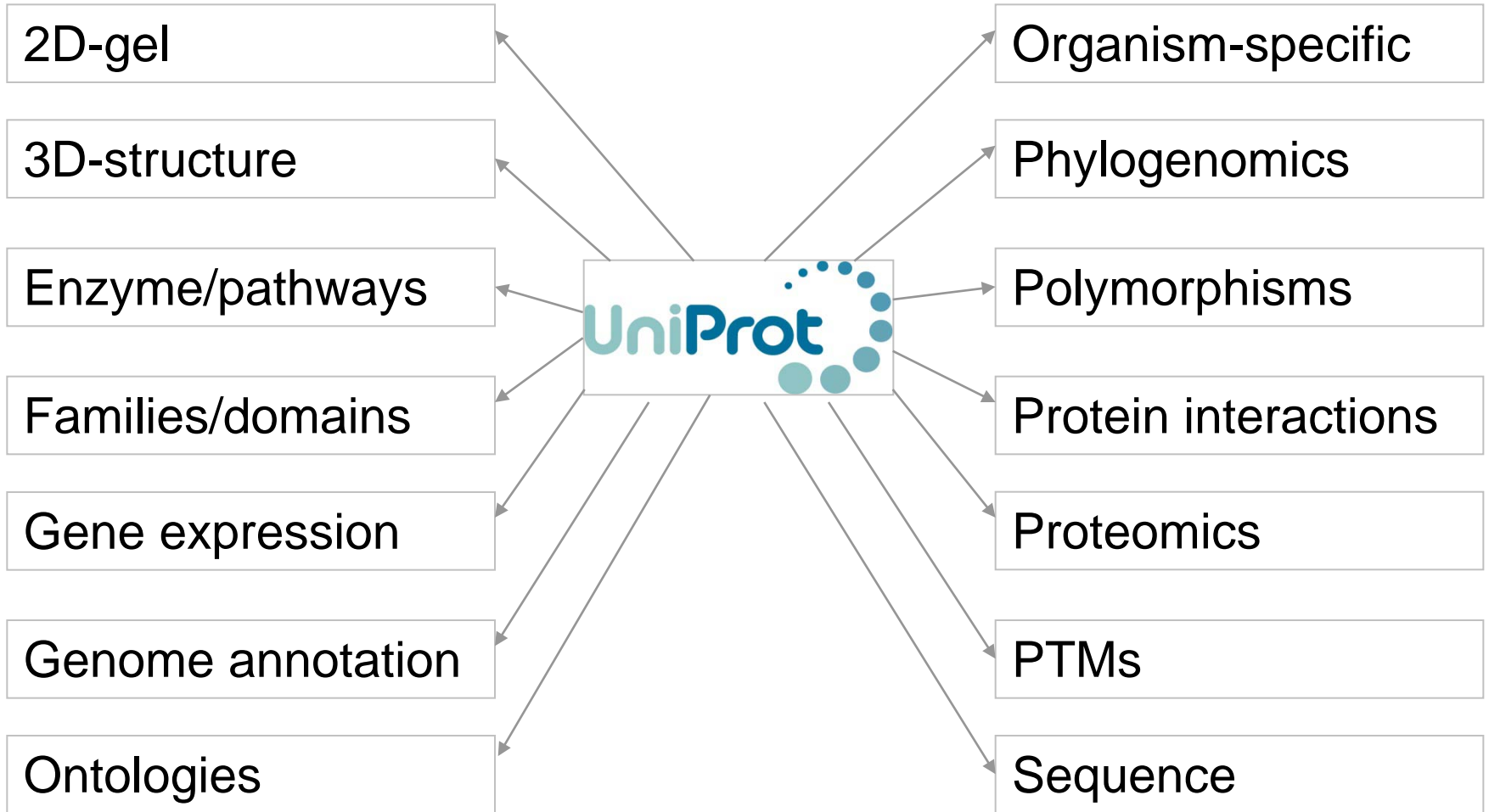
Manual curation



Automatic annotation



Cross-references



Data import

Gene names

Names and origin

Gene names

Name: Msh6 [MGI 1343961](#)

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	



Additional bibliography

The 4.1/ezrin/radixin/moesin domain of the DAL-1/Protein 4.1B tumour suppressor interacts with 14-3-3 proteins.

Yu T., Robb V.A., Singh V., Gutmann D.H., Newsham I.F.

Biochem. J. 365:783-789(2002) [GeneRIF 7529](#) · **Mapped** (22)

Results show that three 14-3-3 isoforms beta gamma and eta are DAL-1/Protein 4.1B-binding proteins. [GeneRIF 7529](#)

Regulation of TSC2 by 14-3-3 binding.

Li Y., Inoki K., Yeung R., Guan K.L.

J. Biol. Chem. 277:44593-44596(2002) [GeneRIF 7529](#) · **Mapped** (22)

TSC2 associates with 14-3-3 in vivo [GeneRIF 7529](#)

GO terms

Gene Ontology (GO)

Biological process

Ras protein signal transduction
Inferred from Experiment. Source: Reactome
activation of pro-apoptotic gene products
Inferred from Experiment. Source: Reactome
cytoplasmic sequestering of protein
Inferred from direct assay. Source: BHF-UCL
negative regulation of protein amino acid dephosphorylation
Inferred from direct assay. Source: BHF-UCL
positive regulation of catalytic activity
Inferred from direct assay. Source: BHF-UCL

Cellular component

cytosol
Inferred from Experiment. Source: Reactome
melanosome
Inferred from electronic annotation. Source: UniProtKB-SubCell
perinuclear region of cytoplasm
Inferred from direct assay. Source: UniProtKB

Molecular function

histone deacetylase binding
Inferred from physical interaction. Source: BHF-UCL
phosphoserine binding
Inferred from physical interaction. Source: BHF-UCL
protein domain specific binding
Inferred from physical interaction. Source: UniProtKB

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	<i>Recommended name:</i> Breast cancer type 2 susceptibility protein <i>Alternative name(s):</i> 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

Search

Blast

Align

Retrieve

ID Mapping

Search in

Query

Protein Knowledgebase (UniProtKB)

Search

Clear

Fields »

CHECKING

Annotation servers loaded:

100%

Cancel

System information:

... Dasty2 finished to resize the graphic.

FILTERING OPTIONS

MANIPULATION OPTIONS (Positional features)



POSITIONAL FEATURES

FEATURE TYPE	METHOD	FEATURE ANNOTATIONS	SERVER NAME	EVIDENCE (Category)
family annotation	InterPro		InterPro-Matches-	inferred from InterPro
family annotation	InterPro		InterPro-Matches-	inferred from InterPro
family annotation	PIRSF		InterPro	inferred from sequence
family annotation	PANTHER		InterPro	inferred from sequence
0-phosphorylated L-serine	NetPhos-2.0		netphos	inferred from electronic
0-phosphorylated L-serine	NetPhos-2.0		cbs total	inferred from electronic
0-phosphorylated L-threonine	NetPhos-2.0		netphos	inferred from electronic
0-phosphorylated L-threonine	NetPhos-2.0		cbs total	inferred from electronic
04'-phosphorylated	NetPhos-2.0		netphos	inferred from electronic
04'-phosphorylated	NetPhos-2.0		cbs total	inferred from electronic
0-(N-acetylaminoalactosyl)-INetOGlyc-3.1	NetOGlyc-3.1		cbs total	inferred from electronic
0-(N-acetylaminoalactosyl)-INetOGlyc-3.1	NetOGlyc-3.1		netoglyc	inferred from electronic
region	PfamB		InterPro	inferred from sequence
region	PfamB		InterPro	inferred from sequence
polypeptide domain	InterPro		InterPro-Matches-	inferred from InterPro
polypeptide domain	InterPro		InterPro-Matches-	inferred from InterPro
polypeptide domain	Pfam		InterPro	inferred from sequence
polypeptide domain	Pfam		InterPro	inferred from sequence
propeptide	ProP-1.0		prop	inferred from electronic
propeptide	ProP-1.0		cbs total	inferred from electronic
polypeptide motif	NetNES-1.1		netnes	inferred from electronic
polypeptide motif	NetNES-1.1		cbs total	inferred from electronic
polypeptide repeat	InterPro		InterPro-Matches-	inferred from InterPro
polypeptide repeat	Pfam		InterPro	inferred from sequence
polypeptide repeat	Pfam		InterPro	inferred from sequence
polypeptide repeat	PS50138		Prosite Features	inferred from reviewed
extramembrane region	CINTHIA		transmem pred	inferred from electronic
non cytoplasmic region	TMHMM-2.0		cbs total	inferred from electronic
non cytoplasmic region	TMHMM-2.0		tmhmm	inferred from electronic
polypeptide conserved region	Everest		everest	inferred from sequence
polypeptide conserved region	Everest		everest	inferred from sequence
polypeptide conserved region	Everest		everest	inferred from sequence
polypeptide conserved region	Everest		everest	inferred from sequence
polypeptide conserved region	Everest		everest	inferred from sequence

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 universally 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-go	GO Annotation of UniProt	2.0	http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot-go	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>

The screenshot displays the UniProt BioMart web interface. At the top, there is a search bar with the text "All Databases" and "Enter Text Here". To the right of the search bar are buttons for "Reset", "Advanced Search", and "Give us feedback". Below the search bar is a navigation menu with tabs for "Databases", "Tools", "EBI Groups", "Training", "Industry", "About Us", and "Help". The "Help" tab is active, and there are sub-links for "Site Index", "RSS", and "Print". Below the navigation menu is a secondary menu with buttons for "UniProt", "New", "Count", "Results", "URL", "XML", "Perl", and "Help". The main content area is divided into two columns. The left column is titled "Dataset" and contains the text "[None selected]". The right column contains a dropdown menu with the following options: "- CHOOSE DATABASE -", "- CHOOSE DATABASE -", "UniProtKB", "INTERPRO (EBI UK)", "PRIDE BioMart", and "ENSEMBL 56 GENES".

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/>

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

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

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