

Protein Interaction Data Sources

Ian Donaldson
Biotechnology Centre of Oslo
University of Oslo



MBV-INF 4410/9410
Monday, November 28th, 2011

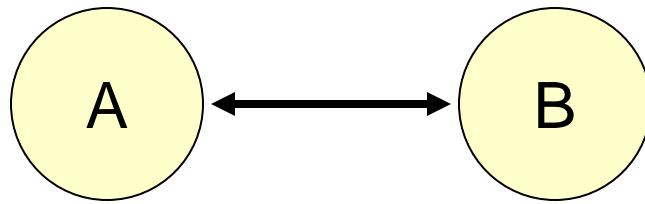
http://donaldson.uio.no/wiki/Protein_Interaction_Resources

<http://irefindex.uio.no>

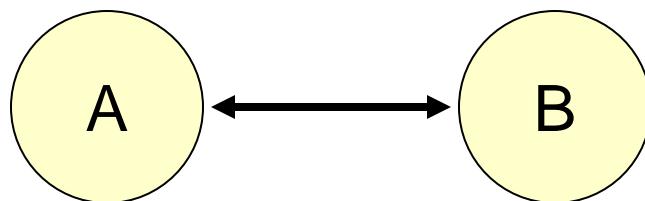
Overview

1. A simple interaction record
2. Experimental methods
3. Representation of interactions
4. Interaction databases
5. Identifiers
6. Text mining sources
7. Predicted interaction sources
8. Consolidated databases
9. Visualization tools
10. Applications

A simple interaction record

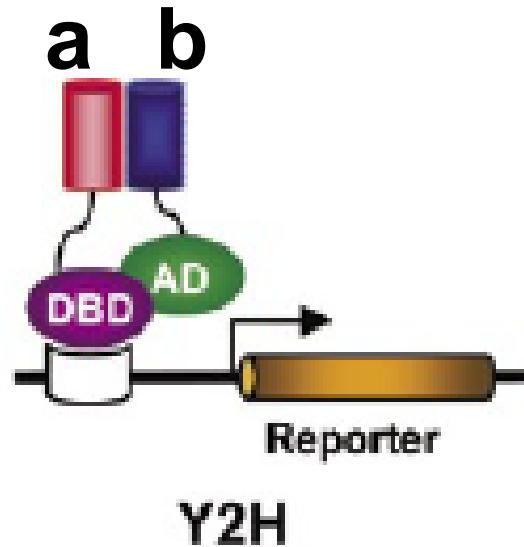


A simple interaction record



1. Short label
2. Type of molecule
3. Database identifier
4. Origin
5. Short label
6. Type of molecule
7. Database identifier
8. Origin
9. Publication reference

Experimental methods – Yeast Two Hybrid



Experimental methods

Two Hybrid family

Advantages

- an "in vivo" binding assay
- easy to automate
- detects binary interactions

Experimental methods

Two Hybrid Family

Limitations

- false positives due to
 - sticky or activating bait or prey
 - overexpression
 - mis-localization (contextual specificity)
 - intermediary proteins
- false negatives due to
 - disruption of interaction by fusion construct
 - absence of intermediary proteins
 - absence of native taxon context

Experimental methods

Affinity purification



- interacting proteins are purified by means of an affinity tag on a "bait" protein
- proteins are detected via MS
- readout of this experiment type is a list of proteins that somehow interact with the bait protein
- note this is fundamentally different than readout of the Y2H family of experiments (more later)

Experimental methods

Affinity purification family

Limitations

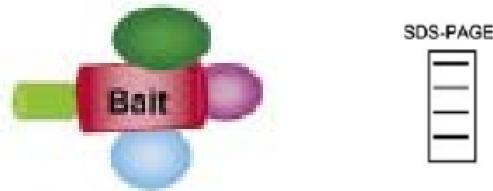
- false positives due to
 - sticky bait or prey
 - mis-identification during MS
 - non-native binding environment
- false negatives due to
 - mis-identification during MS
 - non-native binding environment

Experimental methods

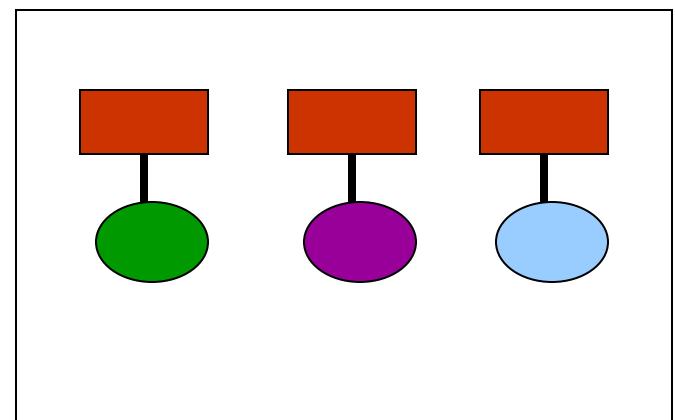
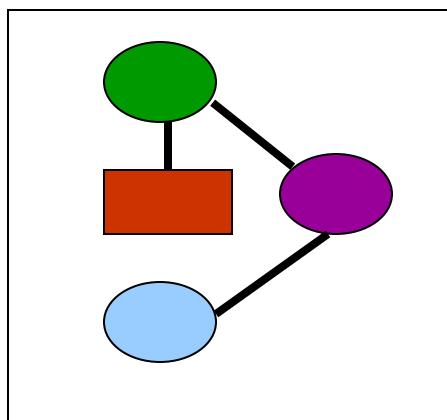
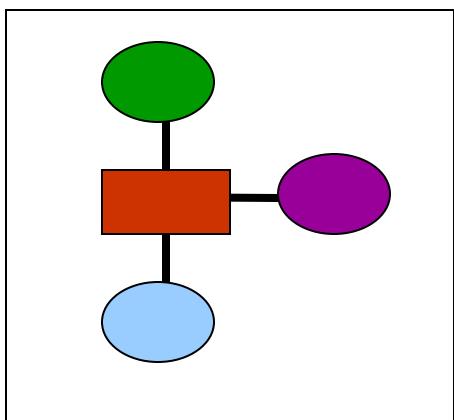
Affinity purification family

Advantages

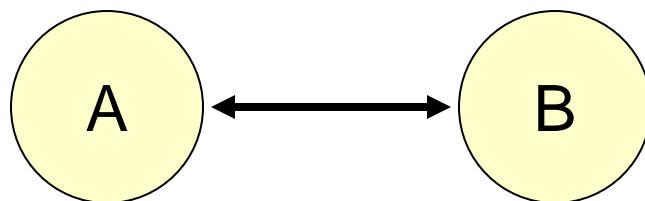
detection of protein complexes but no topology or stoichiometric data



This observation has multiple interpretations:



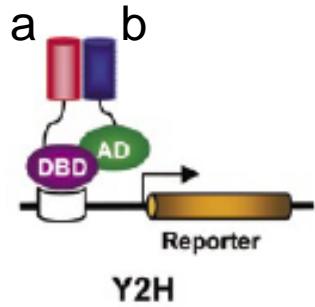
A simple interaction record



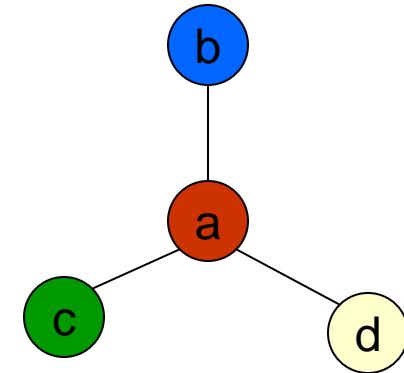
1. Short label
2. Type of molecule
3. Database identifier
4. Origin
5. Short label
6. Type of molecule
7. Database identifier
8. Origin
9. Publication reference

Representation of interaction data

binary data versus associative data

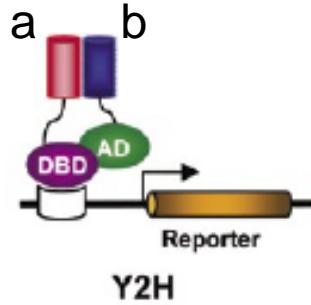


exp	bait	prey
1	a	b
2	a	c
3	a	d

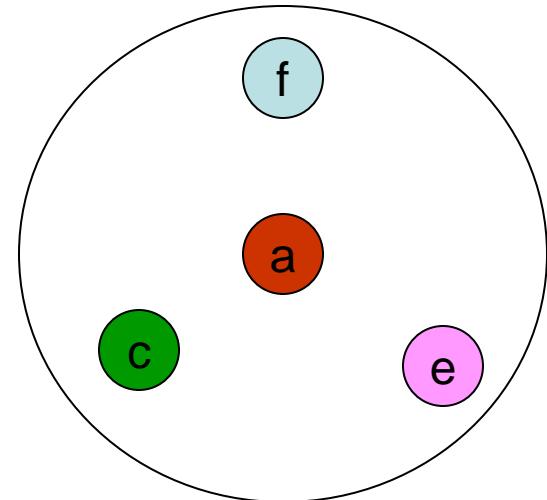
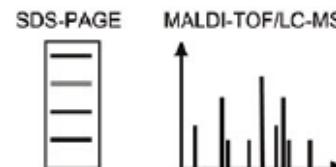
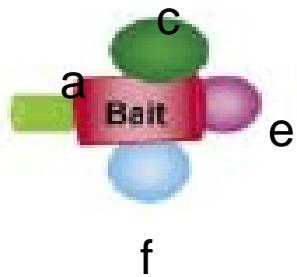
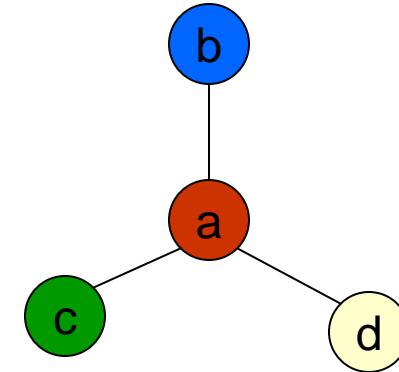


Representation of interaction data

binary data versus associative data

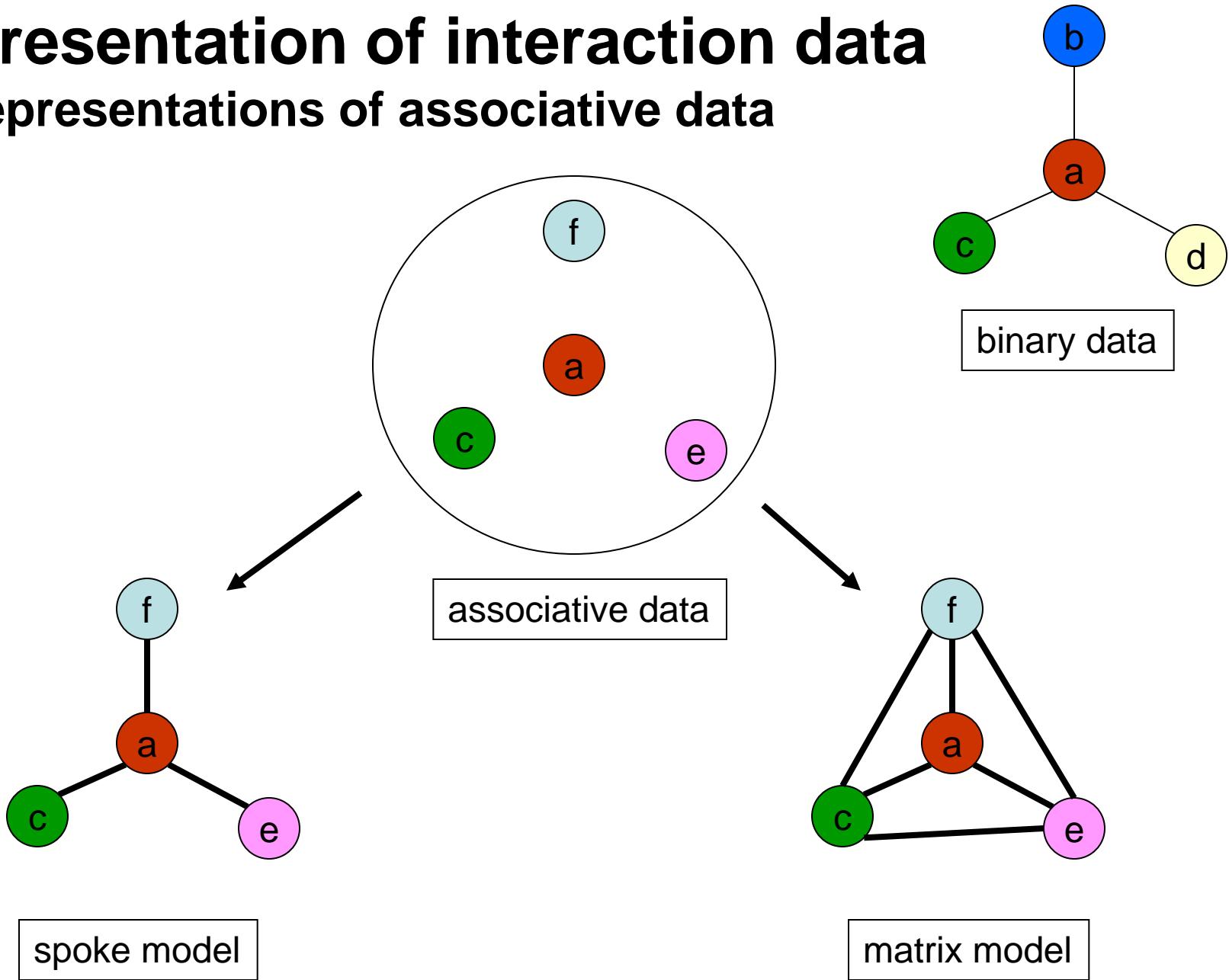


exp	bait	prey
1	a	b
2	a	c
3	a	d

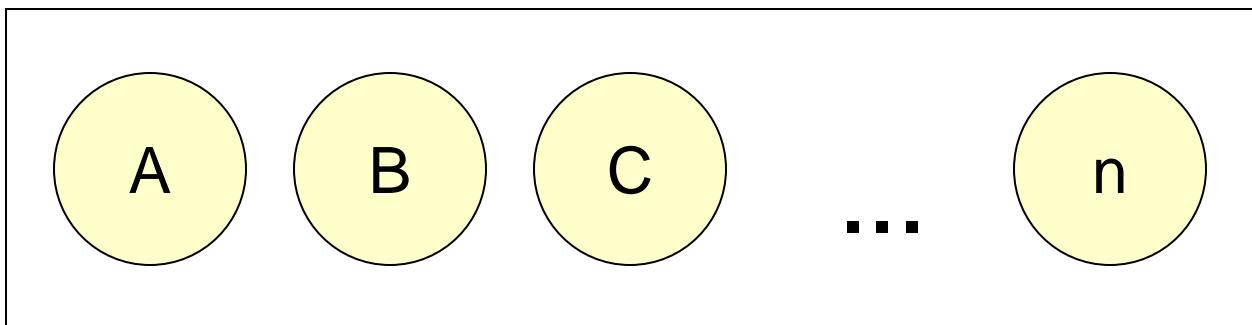


Representation of interaction data

misrepresentations of associative data



A simple interaction record



For each interactor

1. Short label
2. Type of molecule
3. Database identifier
4. Origin

For the interaction

5. Publication reference (method)

More about methods and interaction types that they support

Ontology Lookup Service (OLS) - Mozilla Firefox

<http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=MI>

Ontology Lookup Service (OLS)

EMBL-EBI All Databases Enter Text Here Go Reset Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

OLS - Ontology Lookup Service

MI Ontology Browser

molecular interaction

- interaction detection method
 - + experimental interaction detection
 - + interaction prediction
 - + inference
 - + unspecified method
 - participant identification method
 - feature detection method
 - feature type
- interaction type
 - + genetic interaction
 - + colocalization
 - + association
 - alias type
 - interactor type
 - feature range status
 - experimental preparation
 - cross-reference type
 - database citation
 - experimental role
 - biological role
 - attribute name

Help (hide)

Double-click a term to see its children. The ontology browser is populated dynamically. If there are many children for a given term, there may be a small delay while the browser fetches. Click to highlight a term to see any information associated with it. Hover over a term to see its relation with its immediate parent. Root terms will not display any relational information.

Relations

Term Information

ID: Zoom Name:

Associated information

Highlight a term to view its associated information.

Term Hierarchy

Paths to Root: Child relationships:

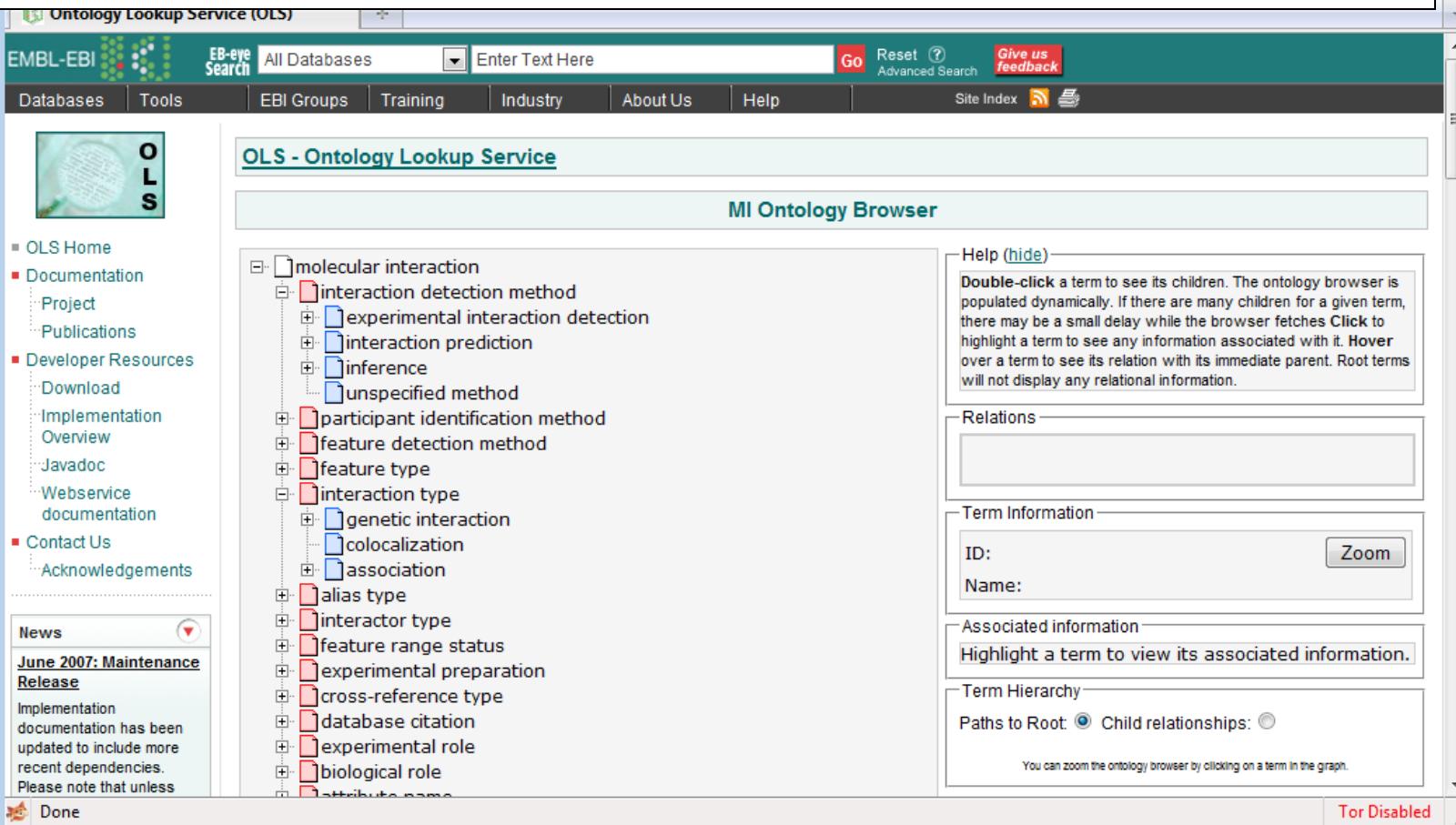
You can zoom the ontology browser by clicking on a term in the graph.

News

June 2007: Maintenance Release

Implementation documentation has been updated to include more recent dependencies. Please note that unless

Done Tor Disabled



The screenshot shows the OLS MI Ontology Browser interface. On the left is a navigation sidebar with links to OLS Home, Documentation, Project, Publications, Developer Resources (Download, Implementation Overview, Javadoc, Webservice documentation), Contact Us (Acknowledgements), and a News section about a June 2007 maintenance release. The main content area displays a hierarchical tree of ontology terms under the root 'molecular interaction'. The tree includes categories like 'interaction detection method', 'interaction prediction', 'inference', 'unspecified method', 'participant identification method', 'feature detection method', 'feature type', 'interaction type' (with subtypes 'genetic interaction', 'colocalization', 'association'), 'alias type', 'interactor type', 'feature range status', 'experimental preparation', 'cross-reference type', 'database citation', 'experimental role', 'biological role', and 'attribute name'. To the right of the tree are several panels: 'Help (hide)', 'Relations', 'Term Information' (with fields for ID and Name), 'Associated information' (with a 'Highlight' link), 'Term Hierarchy' (with options for Paths to Root and Child relationships), and a note about zooming. A status bar at the bottom indicates 'Tor Disabled'.

Interaction databases

BIND	inactive
BioGRID	IMEX (Observer)
CORUM	complexes
DIP	IMEX (Archival)
HPRD	human resource
IntAct	IMEX (Archival)
MATRIX	extracellular matrix
MINT	IMEX (Topical)
MPACT	IMEX (Topical) yeast
MPIDB	IMEX (Observer) prokaryotic
MPPI	mammalian
OPHID/I2D	predicted
Many others	

See <http://www.pathguide.org/> for URLs

Interaction databases

IMEx

International Molecular Exchange Consortium

IMEx partners exchange data and share curation tasks.

Set standards

They also accept interaction data submissions pre and post publication.

Data consolidation and exchange is an active area of research.

See <http://www.imexconsortium.org/>

Interaction databases: IntAct

The screenshot shows the IntAct homepage as it would appear in Mozilla Firefox. The browser's title bar reads "IntAct - Mozilla Firefox". The address bar shows the URL <http://www.ebi.ac.uk/intact/site/index.jsf>. The page content includes:

- IntAct Home**: A large central box containing:
 - Gene name: [BRCA2](#)
 - UniProtKB Ac: [Q06609](#)
 - UniProtKB Id: [dmc1](#)
 - Pubmed Id: [10831611](#)
- IntAct Basic Statistics**: A box listing:
 - 173,621 binary interactions.
 - 54,471 proteins.
 - 9,128 experiments.
 - 1,463 controlled vocabulary terms.
- Introduction**: A brief description of IntAct as a freely available, open source database system for protein interaction data.
- License**: Information about the Apache License version 2 and Creative Commons Attribution License.
- Acknowledgements**: A section at the bottom left.
- Footer**: Includes news items, RSS feeds, and links to upcoming training courses.

Identifiers: EntrezGene

Gene Home - Mozilla Firefox

File Edit View History Bookmarks Diigo Tools Help

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

Most Visited Entrez PubMed Post to CiteULike Library of Medicine and ... The Biotechnology Cen... BBC - bbc.co.uk home... Home page - The Rese...

Diigo Bookmark Highlight Comment Send Message (0) Options

Stumble! I like it! Send to Channels: All Favorites Friends Tools My NCBI [Sign In] [Register]

NCBI Entrez Gene

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Gene for Go Clear

Entrez Gene

Limits Preview/Index History Clipboard Details

Entrez Gene is a searchable database of genes, from RefSeq genomes, and defined by sequence and/or located in the NCBI Map Viewer

News Limit by Chromosomal Region; Sort Options. News archives...

Sample Searches

Find genes by...

- free text
- partial name and multiple species
- chromosome and symbol
- associated sequence accession number
- gene name (symbol)
- publication (PubMed ID)
- Gene Ontology (GO) terms or identifiers

Search text

- human muscular dystrophy
- transporter[title] AND ("Drosophila melanogaster"[orgn] OR "Mus musculus"[orgn])
- (11[chr] OR 2[chr]) AND adh*[sym]
- M11313[accn]
- BRCA1[sym]
- 11331580[PMID]
- "cell adhesion"[GO]
- 10030[GO]

Genes chrom Enzyme

About Entrez Gene

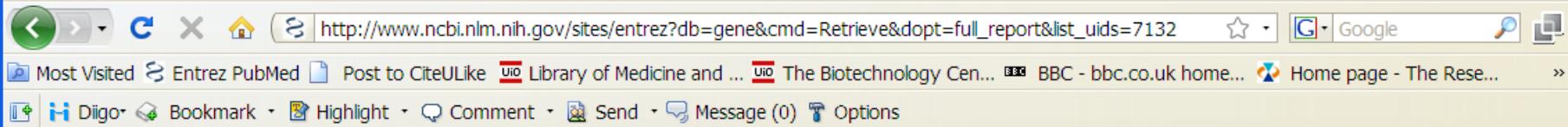
- Entrez Gene: gene-centered information at NCBI, Nucleic Acids Res. 2005 Jan 1;33:D54-8.
- General help on the Entrez search and retrieval system
- NCBI Handbook Chapter on Entrez Gene (download PDF)

- How to add information about function (GeneRIF)
- Correct or update a Gene record
- Correct or update a reference sequence
- Report a publication or GeneRIF error

Done

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

File Edit View History Bookmarks Diigo Tools Help



Stumble! I like it! Send to... Channels: All Favorites Friends Tools

My NCBI
[Sign In] [Register]

Entrez Gene

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Gene for 7132 Go Clear

Limits Preview/Index History Clipboard Details

Display Full Report Show 20 Sort by Relevance Send to

All: 1 Current IP GeneView: 1

GenelD: 7132

superfamily, member 1A [*Homo sapiens*] updated 12-Oct-2008

GeneID: 7132 Summary

Official Symbol **TNFRSF1A**

Official Full Name tumor necrosis

Primary source HGNC:11916

See related Ensembl:ENSG00000067182; HPRD:01861; MIM:191190

Gene type protein coding

RefSeq status REVIEWED

Organism *Homo sapiens*

Lineage *Eukaryota; Metazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Hominoidea; Hominidae; Homo*

Also known as TNF-R-I; p55; p60; TBP1; TNF-R; TNFAR; TNFR1; p55-R; CD120a; TNFR55; TNFR60; TNF-R-I; TNF-R55; MGC19588

Synonyms: TNF-R-I

The protein encoded by this gene is a member of the TNF-receptor superfamily. This protein is one of the major receptors for the tumor necrosis factor-alpha. This receptor can activate NF-kappaB, mediate apoptosis, and function as a regulator of inflammation. Antiapoptotic protein BCL2-associated athanogene 4 (BAG4/SODD) and adaptor proteins TRADD and TRAF2

Entrez Gene Home Table Of Contents

Summary Genomic regions, transcripts... Genomic context Bibliography HIV-1 protein interactions Interactions General gene information General protein information Reference Sequences Related Sequences Additional Links

Links Explain

Order cDNA clone Conserved Domains Genome GEO Profiles HomoloGene Map Viewer Nucleotide EST OMIM PubChem Compound PubChem Substance Full text in PMC Probe

Done

have been shown to interact with this receptor, and thus play regulatory roles in the signal transduction mediated by the receptor. Germline mutations of the extracellular domains of this receptor were found to be associated with the autosomal dominant periodic fever syndrome. The impaired receptor clearance is thought to be a mechanism of the disease. [provided by RefSeq]

Probe
Protein
PubMed
PubMed (OMIM)
PubMed (GeneRIF)
SNP
SNP: Genotype
SNP: GeneView
Taxonomy
UniSTS
AceView
CCDS
Ensembl

Genomic regions, transcripts, and products

(minus strand) Go to reference sequence details

[Try our new Sequence Viewer](#)



RefSeq Protein Accession:
NP_001056

MGL
ModelMaker
PharmGKB
The registry of TNFRSF1A sequence variants
UniGene

Genomic context

chromosome: 12; Location: 12p13.2

[See TNFRSF1A in MapViewer](#)



► [Entrez Gene Info](#)

► [Feedback](#)

► [Subscriptions](#)

Bibliography

Related Articles in PubMed

PubMed links

GeneRIFs: Gene References Into Function

[What's a GeneRIF?](#)

- 1. oxidative stress promotes TNFR receptor (TNFR1- and TNFR2) self-interaction and ligand-independent and enhanced ligand-dependent TNF signaling
- 2. Data show that TNFRSF1A is associated with enhanced TNFR1- and TNFR2-mediated activity

File Edit View History Bookmarks Diigo Tools Help

Back Forward Stop Refresh Address Bar: http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NP_001056.1&dopt=gp

Google

Most Visited Entrez PubMed Post to CiteULike Library of Medicine and ... The Biotechnology Cen... BBC - bbc.co.uk home... Home page - The Rese...

Diigo Bookmark Highlight Comment Send Message (0) Options

Stumble! I like it! Send to Channels: All Favorites Friends Tools

My NCBI
[Sign In] [Register]

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for Go Clear

Display GenPept Show 5 Send to

Range: from begin to end Features: SNP CDD HPRD + Refresh

1: NP_001056. Reports tumor necrosis fa... [gi:450757]

Conserved Domains, Links

[Comment](#) [Features](#) [Sequence](#)

RefSeq Protein Accession: NP_001056

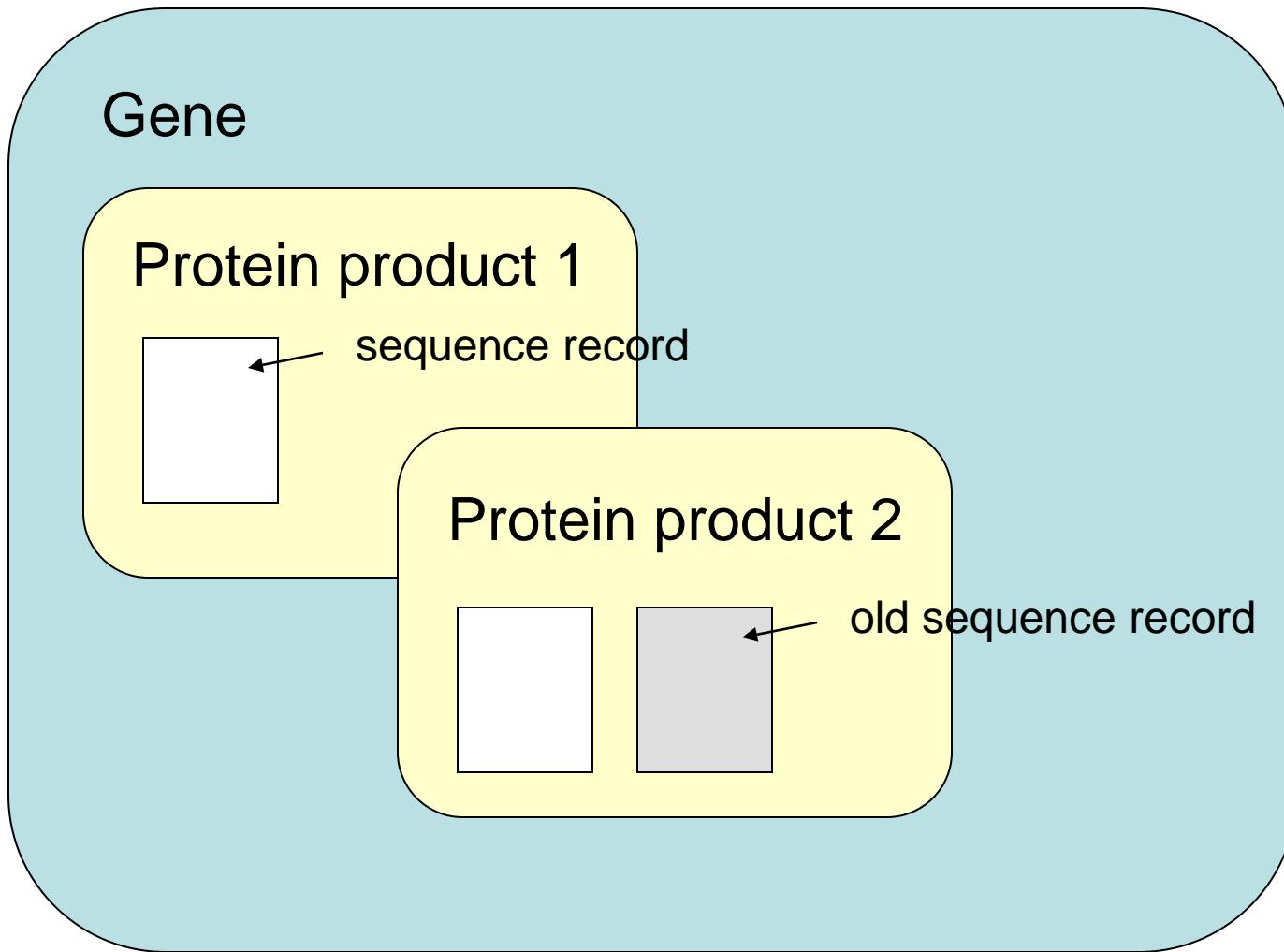
LOCUS NP_001056 455 aa
DEFINITION tumor necrosis factor receptor 1 precursor [Homo sapiens]
ACCESSION NP_001056
VERSION NP_001056.1 GI:4507575
DBSOURCE REFSEQ: accession NM_001065.2
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Catarrhini; Homini
REFERENCE 1 (residues 1 to 455)
AUTHORS Menon,R., Velez,D., Williams,S.M.
TITLE Genetic regulation
receptor concentrations affected by race and protein birth
JOURNAL Hum. Genet. 124 (3), 243-253 (2008)
PUBMED 18807256
REMARK GeneRIF: Observational study of gene-disease association and
gene-environment interaction. (HuGE Navigator)
REFERENCE 2 (residues 1 to 455)
AUTHORS Hosgood,H.D. III, Menashe,I., Shen,M., Yeager,M., Yuenger,J.,
Rajaraman,P., He,X., Chatterjee,N., Caporaso,N.E., Zhu,Y.,

GenInfo Identifier: 4507575

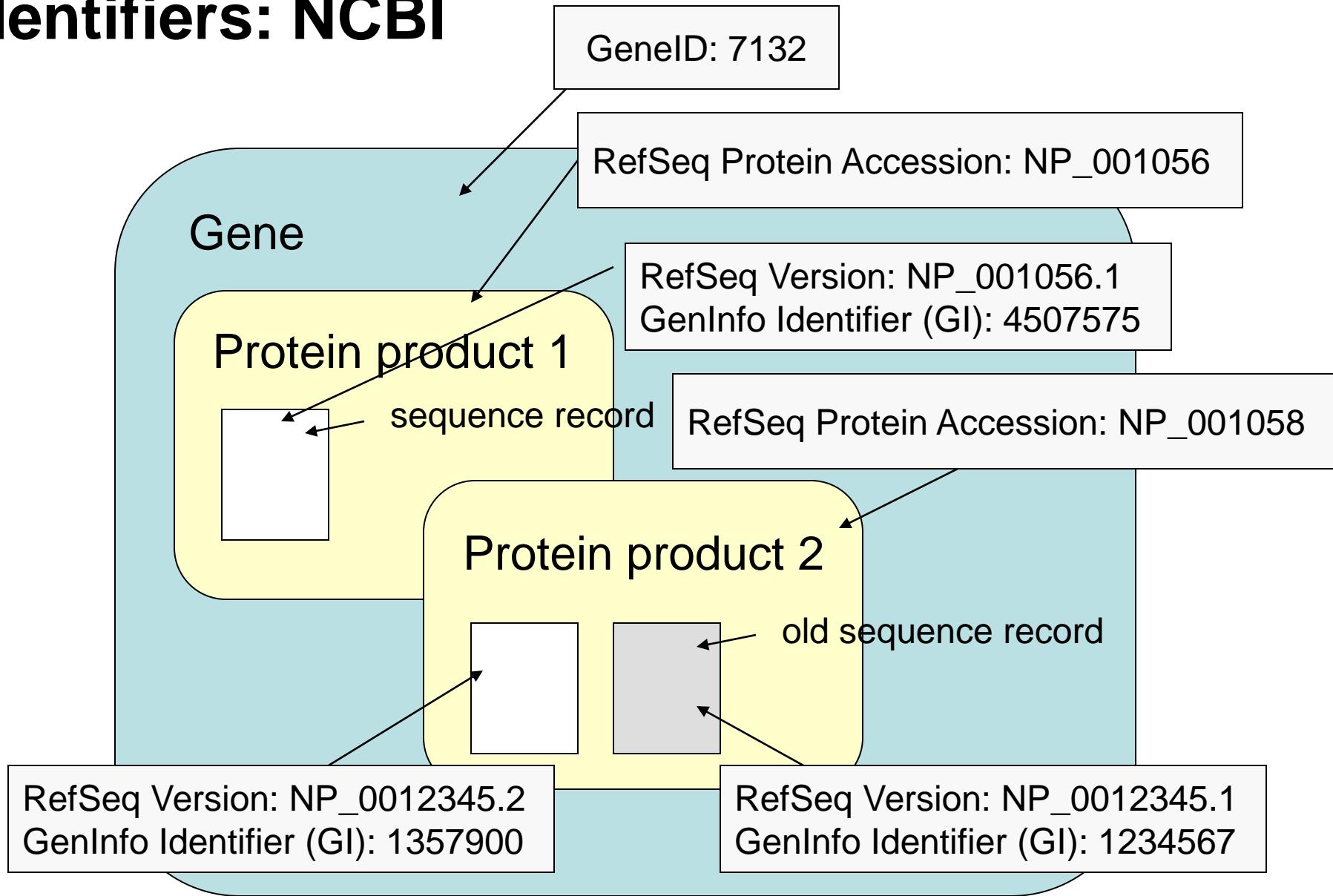
RefSeq Version: NP_001056.1

Done

Identifiers: NCBI



Identifiers: NCBI



Identifiers: UniProtKB

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

UniProt - Mozilla Firefox

File Edit View History Bookmarks Diigo Tools Help

http://www.uniprot.org/

Most Visited Entrez PubMed Post to CiteULike Library of Medicine and ... The Biotechnology Cen... BBC - bbc.co.uk home... Home page - The Rese...

Diigo Bookmark Highlight Comment Send Message (0) Options

Stumble! I like it! Send to Channels: All Favorites Friends Tools

Downloads · Contact · Documentation/Help

Search in Query

Protein Knowledgebase (UniProtKB)

Search Clear Fields »

Search Blast Align Retrieve ID Mapping

WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: <ul style="list-style-type: none">★ Swiss-Prot, which is manually annotated and reviewed.★ TrEMBL, which is automatically annotated and is not reviewed.
UniRef	Sequence clusters, used to speed up similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations, taxonomy, keywords and more.

NEWS

Release 14.3 – Oct 14, 2008
10th anniversary of the Swiss Institute of Bioinformatics

Statistics for UniProtKB: Swiss-Prot · TrEMBL
Forthcoming changes
News archives

SITE TOUR



Learn how to make best use of the tools and data on this site.

Done

DNA repair protein RAD51 homolog 1 - Homo sapiens (Human) - Mozilla Firefox

File Edit View History Bookmarks Digo Tools Help

http://www.uniprot.org/uniprot/Q06609 uniprot

Most Visited Entrez PubMed Post to CiteULike Library of Medicine and ... The Biotechnology Cen... BBC - bbc.co.uk home... Home page - The Rese... >

Stumble! I like it! Send to Channels: All Favorites Friends Tools

DNA repair protein RAD51... Entrez Gene: RAD51 RAD51 h...

UniProtKB Entry: RAD51_HUMAN

UniProtKB Accession: Q06609

Reviewed, UniProtKB/Swiss-Protein Q06609 (RAD51_HUMAN)

Last modified September 23, 2008. Version 100. History

Clusters with 100%, 90%, 50% identity | Documents | PDF/XML | GFF | FASTA

Names and origin · Protein attributes · General annotation (Comments) · Ontologies · Binary interactions · Alternative products · Sequence annotation (Features) · Sequences · References · Cross-references · Entry information · Relevant documents

Names and origin Hide | Top

Protein names Recommended name: DNA repair protein RAD51
Short name=hRAD51
Short name=HsRAD51

Gene names Name: RAD51
Synonyms: RAD51A, RECA

Organism Homo sapiens (Human)

Taxonomic identifier 9606 [NCBI]

Taxonomic lineage Eukaryota > Metazoa > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo

Gene Name: RAD51

Synonyms: RAD51A

Taxon id: 9606 (human)

File Edit View Bookmarks Diigo Tools Help

http://www.uniprot.org/uniprot/Q06609

RSS Google

Most Visited Entrez PubMed Post to CiteULike Library of Medicine and ... The Biotechnology Cen... BBC BBC - bbc.co.uk home... Home page - The Rese... Wikipedia

Diigo Bookmark Highlight Comment Send Message (0) Options

Stumble! Like it! Send to Channels: All Favorites Friends Tools

Binary interactions

Hide | Top

Binary interactions from IntAct

With	Entry		
itself		1	EBI-297202,EBI-297202
BRCA2	P51587	7	EBI-297202,EBI-79792
RAD51AP1	Q96B01-2	2	EBI-297202,EBI-1178743
RAD51AP1	Q96B01-3	3	EBI-297202,EBI-1178748
rhp51	P36601	2	EBI-297202,EBI-926960 From a different organism.
TP53	P04637	1	EBI-297202,EBI-366083

Alternative products

Hide | Top

This entry describes 2 isoforms produced by **alternative splicing**. [Align] [Select]**Isoform 1** (identifier: Q06609-1)*This isoform has been chosen as the canonical version of the entry.***Isoform identifiers***The sequence of this isoform refers to it. This is also the sequence that appears in the downloadable files.***Isoform 2** (identifier: Q06609-2)*The sequence of this isoform differs from the canonical sequence as follows:*

77-173: Missing.

Notes: No experimental confirmation available.

Find: dmc1 Next Previous Highlight all Match case Reached end of page, continued from top

Done

DNA repair protein RAD51 homolog 1 - Homo sapiens (Human) - Mozilla Firefox

File Edit View History Bookmarks Diigo Tools Help

http://www.uniprot.org/uniprot/Q06609

Most Visited Entrez PubMed Post to CiteULike Library of Medicine and ... The Biotechnology Cen... BBC BBC - bbc.co.uk home... Home page - The Rese... Wikipedia

Diigo Bookmark Highlight Comment Send Message (0) Options

Stumble! I like it! Send to Channels: All Favorites Friends Tools

Other Resources

SOURCE	Search...
ProtoNet	Search...

Entry information Hide | Top

Entry name	RAD51_HUMAN
Accession	Primary (citable) accession number: Q06609 Secondary accession number(s): Q6FHX9, Q9BV60
Entry history	Integrated into UniProtKB/Swiss-Prot: June 1, 1994 Last sequence update: June 1, 1994 Last modified: September 1, 1994 This is version 100 of the entry and version 1 of the sequence. [Complete history]
Entry status	Reviewed (UniProtKB/Swiss-Prot)
Annotation project	HPI (Human Proteome Initiative)

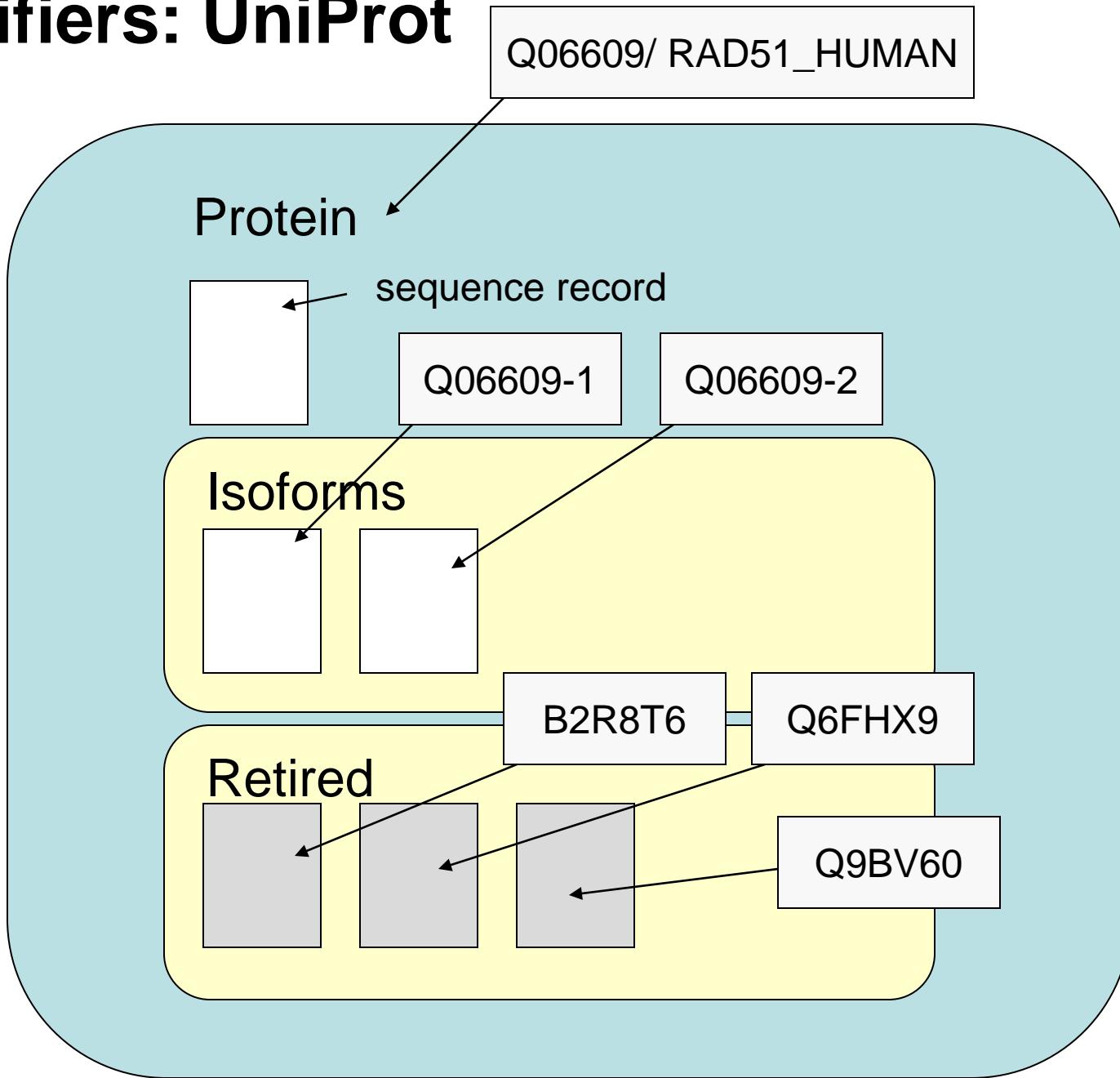
Relevant documents Hide | Top

- Human chromosome 15
- Human chromosome 15: entries, gene names and cross-references to MIM
- Human entries with polymorphisms or disease mutations
- List of human entries with polymorphisms or disease mutations
- Human polymorphisms and disease mutations
- Index of human polymorphisms and disease mutations

Find: dmc1 Next Previous Highlight all Match case Reached end of page, continued from top

Done

Identifiers: UniProt



Common Identifiers

Gene

Ensembl ENSG00000139618

Entrez Gene 675

Unigene Hs.34012

RNA transcript

GenBank BC026160.1

RefSeq NM_000059

Ensembl ENST00000380152

Protein

Ensembl ENSP00000369497

RefSeq NP_000050.2

UniProt BRCA2_HUMAN or

A1YBP1_HUMAN

IPI IPI00412408.1

EMBL AF309413

PDB 1MIU

Species-specific

HUGO HGNC BRCA2

MGI MGI:109337

RGD 2219

ZFIN ZDB-GENE-060510-3

FlyBase CG9097

WormBase WBGene00002299 or ZK1067.1

SGD S000002187 or YDL029W

Annotations

InterPro IPR015252

OMIM 600185

Pfam PF09104

Gene Ontology GO:0000724

SNPs rs28897757

Experimental Platform

Affymetrix 208368_3p_s_at

Agilent A_23_P99452

CodeLink GE60169

Illumina GI_4502450-S

Red = Recommended

UniProt ID Mapping Service

The screenshot shows the UniProt homepage in a Mozilla Firefox browser window. The URL in the address bar is <http://www.uniprot.org/>. The browser interface includes a menu bar (File, Edit, View, History, Bookmarks, Tools, Help), a toolbar with various icons, and a tab bar showing multiple open tabs. The main content area features a search bar with dropdown options for "Search in" (Protein Knowledgebase (UniProtKB)) and "Query", and buttons for "Search", "Clear", and "Fields ». Below the search bar are buttons for "Search", "Blast", "Align", "Retrieve", and "ID Mapping". A red arrow points from the bottom right towards the "ID Mapping" button. To the left of the search bar, there's a "WELCOME" section with a mission statement and a "What we provide" table. To the right, there are "NEWS" and "SITE TOUR" sections.

WELCOME

The mission of **UniProt** is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: <ul style="list-style-type: none">★ Swiss-Prot, which is manually annotated and reviewed.★ TrEMBL, which is automatically annotated and is not reviewed.
UniRef	Sequence clusters, used to speed up similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations , taxonomy , keywords and more.

NEWS

UniProt release 15.7 – Sep 1, 2009

Formyl peptide receptors: the missing link between olfaction and immune system · Cross-references to STRING

› Statistics for UniProtKB:
 Swiss-Prot · TrEMBL
› Forthcoming changes
› News archives

SITE TOUR

Learn how to make best use of the tools and data on this site.

UniProt ID Mapping Service

The screenshot shows the UniProt ID Mapping Service interface within a Mozilla Firefox browser window. The main content area displays the 'Identifiers' mapping tool. On the left, under 'Identifiers', the identifier 'YIL062C' is listed. The 'From' dropdown is set to 'UniProtKB AC/ID' and the 'To' dropdown is set to 'Entrez Gene (GeneID)'. Below these are buttons for 'Map', 'Swap', and 'Clear'. To the right of the mapping tool is a box titled 'Database identifier mapping tips' containing instructions and a 'More...' link. At the bottom of the mapping tool section are buttons for 'Search', 'Blast', 'Align', 'Retrieve', and 'ID Mapping'. Below this, the 'WELCOME' section introduces the mission of UniProt, stating it provides a comprehensive, high-quality, and freely accessible resource of protein sequence and functional information. The 'What we provide' section highlights the UniProtKB knowledgebase, which consists of two sections: Swiss-Prot (manually annotated and reviewed) and TrEMBL (automatically annotated). The right sidebar features a 'NEWS' section with a link to 'UniProt release 15.7 – Sep 1, 2009' and a 'SITE TOUR' section.

UniProt - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.uniprot.org/ uniprot

Most Visited Getting Started Latest Headlines

Stumble! I like it! All Share Info Favorites Friends Tools

Bioinformatics course - Donaldson ... UniProt

Downloads · Contact · Documentation/Help

Identifiers

YIL062C

From: UniProtKB AC/ID
To: Entrez Gene (GeneID)
or Browse...

Map Swap Clear

Database identifier mapping tips

To map identifiers to or from UniProtKB:

- enter identifiers, e.g.: ITIA 1FNS
- select a source database, e.g.: PDB
- or select a target database, e.g.: UniProtKB

More...

Search Blast Align Retrieve ID Mapping

WELCOME

The mission of [UniProt](#) is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: <ul style="list-style-type: none">★ Swiss-Prot, which is manually annotated and reviewed.★ TrEMBL, which is automatically annotated and is
-----------	--

NEWS

[UniProt release 15.7 – Sep 1, 2009](#)

[Formyl peptide receptors: the missing link between olfaction and immune system](#) · Cross-references to STRING

[Statistics for UniProtKB: Swiss-Prot · TrEMBL](#)

[Forthcoming changes](#)

[News archives](#)

SITE TOUR

Ensembl BioMart

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

The screenshot shows the Ensembl BioMart interface running in Mozilla Firefox. The browser's address bar displays the URL <http://www.biomart.org/biomart/martview/02e2f0a0dabf3992cb35496f164dbaad>. The main content area is the BioMart - MartView page. At the top, there is a navigation bar with links to HOME, MARTVIEW (which is active), MARTSERVICE, DOCS, CONTACT, NEWS, and CREDITS. Below this is a toolbar with buttons for New, Count, and Results. A menu bar includes File, Edit, View, History, Bookmarks, Tools, and Help. On the left, a sidebar titled 'Dataset' contains sections for Features, Filters (selected), Attributes (Feature Symbol, Feature Name, GO ID, GO Term), and another Dataset section with [None Selected]. The main query form is titled 'Please restrict your query using criteria below'. It features three expandable sections: 'Identifiers' (set to MGI ID), 'Feature Name' (empty input field), and 'Feature Type' (a dropdown menu listing Gene, DNA Segment, Cytogenetic Marker, QTL, and RNA). A 'Done' button is located at the bottom left of the query form.

ID Mapping Challenges

- Avoid errors: map IDs correctly
- Gene name ambiguity – not a good ID
 - e.g. FLJ92943, LFS1, TRP53, p53
 - Better to use the standard gene symbol: TP53
- Excel error-introduction
 - OCT4 is changed to October-4
- Problems reaching 100% coverage
 - E.g. due to version issues
 - Use multiple sources to increase coverage

Zeeberg BR et al. Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics BMC Bioinformatics. 2004 Jun 23;5:80

ID Mapping Challenges

- Spot-test any ID mapping service you use.
- Check samples from first, last and middle of your list of identifiers to be converted.
- Ask for help if you are uncertain.

Identifiers: Why can't I find my protein

Are you using a sequence record identifier?

Is the identifier retired?

Are you using the identifier for the correct organism?

Have you tried searching for identifiers for all products of the gene?

Try different identifier systems (Entrez Gene, UniProt)

Remember that only 10% to 50% of all proteins have interaction information.

You always have recourse to text-mining and predicted interactions.

Overview

1. A simple interaction record
2. Experimental methods
3. Representation of interactions
4. Interaction databases
5. Identifiers
6. Text mining sources
7. Predicted interaction sources
8. Consolidated databases
9. Visualization tools
10. Applications

Text mining resources

Most are based on co-occurrence of gene name mentions.

Start at http://zope.bioinfo.cnio.es/bionlp_tools

iHOP	<u>http://www.ihop-net.org/UniPub/iHOP/</u>
EBIMed	<u>http://www.ebi.ac.uk/Rebholz-srv/ebimed/index.jsp</u>
FACTA	<u>http://text0.mib.man.ac.uk/software/facta/</u>
PubGene	<u>http://www.pubgene.org/</u>
InfoPubMed	<u>https://www-tsujii.is.s.u-tokyo.ac.jp/info-pubmed/</u>
Chillibot	<u>http://www.chilibot.net/</u>
MedEvi	<u>http://www.ebi.ac.uk/tc-test/textmining/medevi/</u>
FABLE	<u>http://fable.chop.edu/overview.jsp</u>
BCMS	<u>http://bcms.bioinfo.cnio.es/</u>
BioText	<u>http://biosearch.berkeley.edu/</u>

Text mining resources: iHOP

iHOP - Information Hyperlinked over Proteins - Mozilla Firefox

File Edit View History Bookmarks Diigo Tools Help

Back Forward Stop Refresh Address bar: http://www.ihop-net.org/UniPub/iHOP/index.html?field=all&search=RAD51&organism_id=0 Favorites biotext Search

Most Visited Entrez PubMed Post to CiteULike Library of Medicine and ... The Biotechnology Cen... BBC - bbc.co.uk home... Home page - The Rese... Wikipedia

Diigo Bookmark Highlight Comment Send Message (0) Options

Stumble! I like it! Send to Channels: All Favorites Friends Tools

iHOP - Info... EBIMed FACTA PubGene - Find... Chilibot: finding... MedEvi - a per... FABLE - Overv... BioCreative Met... BioText Search ...

iHOP
information hyperlinked over proteins

Search Gene RAD51

Gene Model
Developer's Zone new
How to cite iHOP

Contact Links Help

WILHELMUS KETTERING CANCER CENTER
ESTABLISHED 1898

cocoon

Concept & Implementation by Robert Hoffmann

Symbol Name Synonym/ DB-reference Organism Results

RAD51 RAD51 homolog (RecA homolog, E. coli)
(S. cerevisiae)

Rad51 RAD51 homolog (S. cerevisiae)

RAD51 Strand exchange protein, forms a helical filament with DNA that searches for ...

rad51 RAD51 homolog (RecA homolog, E. coli)
(S. cerevisiae)

RAD51 LeRAD51

rad51 rad51 protein

RAD51 RAD51 homolog (RecA homolog, E. coli)
(S. cerevisiae)

RAD51 RAD51 homolog (RecA homolog, E. coli)
(S. cerevisiae)

rad51 DNA repair protein Rad51 homolog

rad51 DNA repair protein

RAD51L1 RAD51-like 1 (S. cerevisiae)

RAD51C RAD51 homolog C (S. cerevisiae)

RAD51L3 RAD51-like 3 (S. cerevisiae)

RAD51AP1 RAD51 associated protein 1

Rad51l3 RAD51-like 3 (S. cerevisiae)

RAD51L4 RAD51-like 4 (S. cerevisiae)

Life cycles of successful genes

Homo sapiens

Mus musculus

Saccharomyces cerevisiae

Danio rerio

Lycopersicon esculentum

Xenopus laevis

Gallus gallus

Canis familiaris

Guillardia theta

Giardia lamblia ATCC 50803

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Mus musculus

Interaction information for this gene.

Find: dmc1 Next Previous Highlight all Match case Reached end of page, continued from top

javascript:if(window.name=="") { window.location.href='./nil'; } else { doaction('IHOP', 91604, 1); }

Navigation icons: Back, Forward, Stop, Refresh, Home, etc.

Text mining resources: PubGene

Network based on co-occurrence in article for the "RAD51" Gene and Proteins in *Homo sapiens* - Mozilla Firefox

File Edit View History Bookmarks Digo Tools Help

http://www.pubgene.org/ biotext

Most Visited Entrez PubMed Post to CiteULike Library of Medicine and ... The Biotechnology Cen... BBC - bbc.co.uk home... Home page - The Rese... Wikipedia Google Translate

Bookmark Highlight Comment Send Message (0) Options

Stumble! I like it! Send to... Channels: All Favorites Friends Tools

PubGene - Find ... Chillbot: finding ... MedEvi - a perm... FABLE - Overview BioCreative Met... BioText Search ... RAD51 - Google... RAD51 - Google... Network base...

Bio Networks **Bio Associations** **Sequence Homology** **My Workspace Login**

Browse literature or sequence neighbours.

Organism: Homo sapiens Select one or "all organisms"
Gene / Protein *: RAD51 Terms separated by comma, eg. *BRCA1, LEP*
and/or
Biological term: One term, eg. *Alzheimer Disease*.**Supported categories**
Network displays genes/proteins important for the biological term

Submit [Advanced Options](#) | [Network Parameters](#)

Network based on co-occurrence in article for the "RAD51" Gene and Proteins in *Homo sapiens*

[Send selected terms to](#) [Choose PubGene Tool](#)

[Information about 'RAD51'](#)

[Key information](#) [Related terms](#)

Property	Value
Term	RAD51
Name	RAD51 homolog (Re...
Organism	<i>Homo sapiens</i>
Article	4905
Term T...	PubGene Gene/Prot...
External...	EG
Chrom...	
Alt. Sym...	BRCC5
Alt. Sym...	HRAD51
Alt. Sym...	HsRad51
Alt. Sym...	HsT16930
Alt. Sym...	RAD51A
Alt. Sym...	RECA

[Network view](#) [Table view](#)

The network graph displays interactions between various genes/proteins. Nodes include RAD51, POLG2, ATRX, POLA1, BRCA2, BRCA1, TP53, DIAH3, STK32C, and ATP5E. RAD51 is at the center, connected to POLG2, ATRX, POLA1, BRCA2, DIAH3, and TP53. BRCA2 is connected to TP53 and DIAH3. TP53 is connected to STK32C. DIAH3 is connected to ATP5E.

Find: dmc1 Next Previous Highlight all Match case Reached end of page, continued from top

Applet pubgenegraph.GraphController started

Overview

1. A simple interaction record
2. Experimental methods
3. Representation of interactions
4. Interaction databases
5. Identifiers
6. Text mining sources
7. Predicted interaction sources
8. Consolidated databases
9. Visualization tools
10. Applications

Predicted interaction resources

STRING <http://string.embl.de/>

OPHID/I2D <http://ophid.utoronto.ca/ophidv2.201/>

FunCoup <http://FunCoup.sbc.su.se>

Finding interaction information

The take home message:

There is no single place to collect interaction information.

Users are placed in a "go-fish" position and are responsible for distinguishing and sorting through multiple data sources.

There are a few resources that are focusing on providing an integrated data resource...

Overview

1. A simple interaction record
2. Experimental methods
3. Representation of interactions
4. Interaction databases
5. Identifiers
6. Text mining sources
7. Predicted interaction sources
- 8. Consolidated databases**
9. Visualization tools
10. Applications

Consolidated databases

APID2NET

<http://bioinfow.dep.usal.es/apid/apid2net.html>

Pathway Commons

<http://www.pathwaycommons.org/pc/>

MiMI

<http://mimi.ncibi.org/MimiWeb/main-page.jsp>

iRefIndex

<http://irefindex.uio.no>

Finding interaction information

The take home message:

There is no single place to collect interaction information.

Users are placed in a "go-fish" position and are responsible for distinguishing and sorting through multiple data sources.

There are a few resources that are focusing on providing an integrated data resource...

Overview

1. A simple interaction record
2. Experimental methods
3. Representation of interactions
4. Interaction databases
5. Identifiers
6. Text mining sources
7. Predicted interaction sources
8. Consolidated databases
9. Visualization tools
10. Applications

Visualization Tools

Cytoscape

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

VisAnt

<http://visant.bu.edu/>

There are over 30 visualization tools listed in a recent review

Suderman M, Hallett M: Tools for visually exploring biological networks.
Bioinformatics 2007, 23(20):2651-2659.

Cytoscape will be covered in a following talk.

Conclusions

Interaction data is an important bioinformatics data source.

Interaction detection techniques have evolved past the simple Y2H and coimmunoprecipitation methods.

Finding data is still a non-trivial task.

There is no de facto source of data and users are faced with doing their own collection and curation.

Data integration and accessibility is an ongoing area of research.