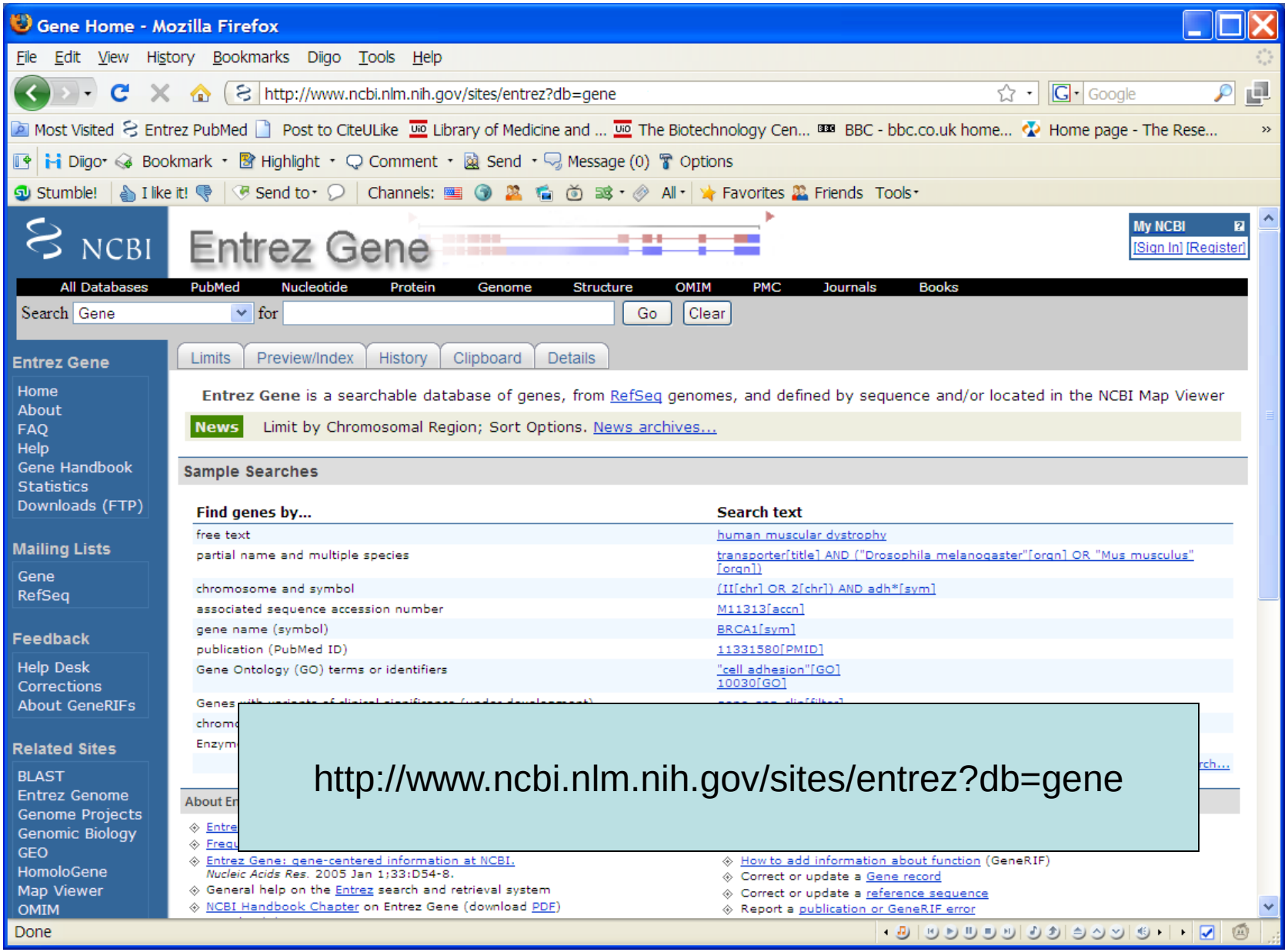
# Database lecture
# Working with common db identifiers

Ian Donaldson
Biotechnology Centre of Oslo
University of Oslo

MBV-INF 4410/9410

# Identifiers: EntrezGene




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

Entrez Gene: TNFRSF1A tumor necrosis factor receptor superfamily, member 1A [ Homo sapiens ] - Mozilla Firefox

File Edit View History Bookmarks Diigo Tools Help

http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Retrieve&dopt=full_report&list_uids=7132

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

NCBI Entrez Gene

My NCBI [Sign In] [Register]

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 | Curre... | ...IP GeneView: 1

GeneID: 7132

1: TNFRSF1... superfamily, member 1A [ *Homo sapiens* ]

updated 12-Oct-2008

Summary

| | |
|---|---|
| Official Symbol | TNFRSF1A ... ed by HGNC |
| Official Full Name | tumor necrosis ... provided by HGNC |
| Primary source | HGNC:11916 |
| See related | Ensembl:ENSG00000067182; HPRD:01861; MIM:191190 |
| Gene type | protein coding |
| RefSeq status | REVIEWED |
| Organism | ... |
| Lineage | ...orata; Euteleostomi; Mammalia; Eutheria; ...ini; Hominidae; Homo |
| Also known as | FPF; p55; p60; TBP1; TNF-R; TNFAR; TNFR1; p55-R; CD120a; TNFR55; TNFR60; TNF-R-I; TNF-R55; MGC19588 |
| Summary | 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 |

GeneID: 7132

Official Symbol: TNFRSF1A
aka Gene Name

Synonyms: TNF-R-I

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

Entrez Gene: TNFRSF1A tumor necrosis factor receptor superfamily, member 1A [ Homo sapiens ] - Mozilla Firefox

File Edit View History Bookmarks Diigo Tools Help

http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Retrieve&dopt=full_report&list_uids=7132

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

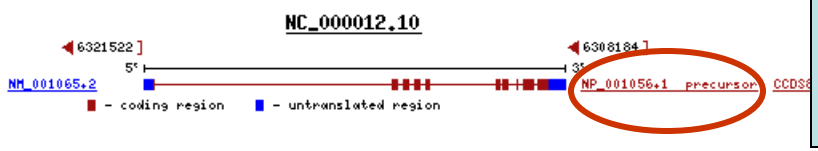
Stumble! | I like it! | Send to | Channels: | All | Favorites | Friends | Tools

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]

## Genomic regions, transcripts, and products

(minus strand) Go to reference sequence details

Try our new Sequence Viewer

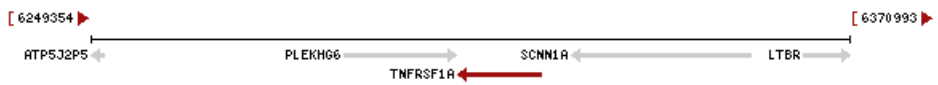



RefSeq Protein Accession: NP_001056

## Genomic context

chromosome: 12; Location: 12p13.2

See TNFRSF1A in MapViewer




## 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

Probe
Protein
PubMed
PubMed (OMIM)
PubMed (GeneRIF)
SNP
SNP: Genotype
SNP: GeneView
Taxonomy
UniSTS
AceView
CCDS
MGC
ModelMaker
PharmGKB
The registry of TNFRSF1A sequence variants
UniGene

Entrez Gene Info

Feedback

Subscriptions

Done

NCBI Sequence Viewer v2.0 - Mozilla Firefox

File Edit View History Bookmarks Diigo Tools Help

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NP_001056.1&dopt=gp

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

NCBI Entrez Protein

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

```
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
  AUTHORS   Menon,R., Velez,D.
            Williams,S.M.
  TITLE     Genetic regulation
            receptor concentrations affected by race and preterm 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.,
```

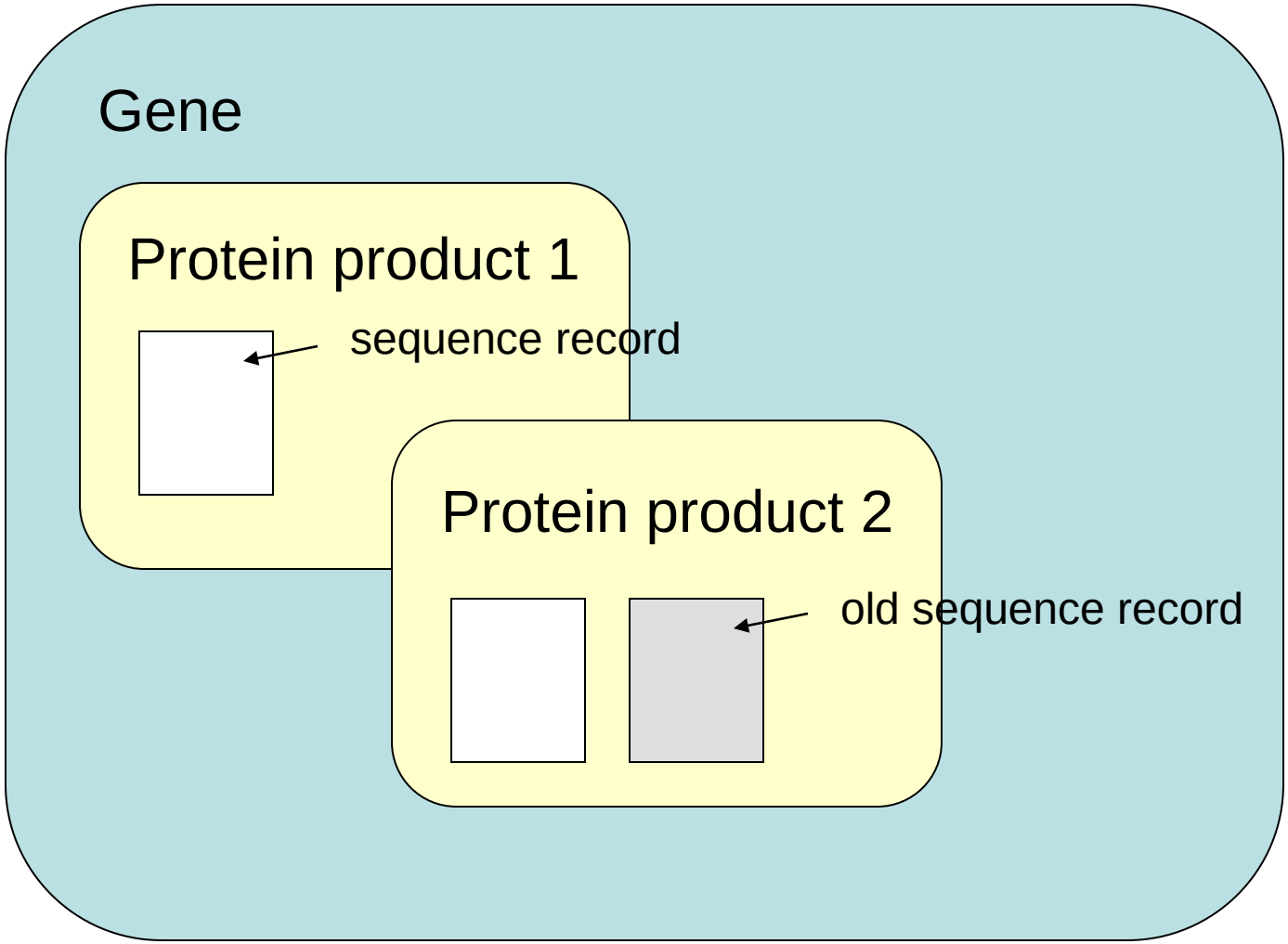
RefSeq Protein Accession: NP_001056

GenInfo Identifier: 4507575

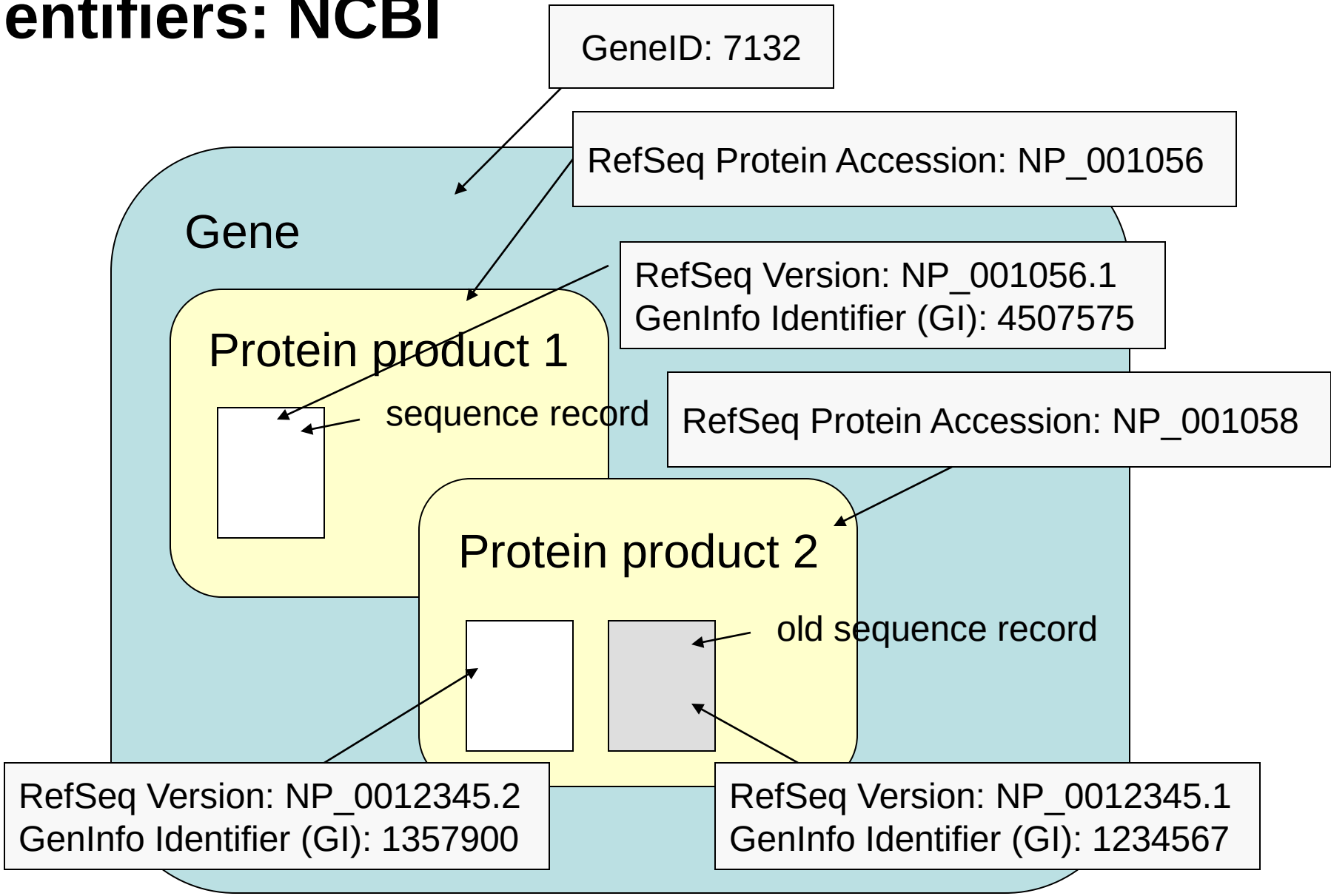
RefSeq Version: NP_001056.1

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

UniProt

Downloads · Contact · Documentation/Help

Search in: Protein Knowledgebase (UniProtKB)

Query: 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:<br>Swiss-Prot, which is manually annotated and reviewed.<br>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 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.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

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

UniProt › UniProtKB Downloads · Contact · Documentation/Help

Search in: Protein Knowledgebase (UniProtKB) | Query

UniProtKB Entry: RAD51_HUMAN

Reviewed, UniProtKB/Swiss-Prot Q06609 (RAD51_HUMAN)

Last modified September 23, 2008. Version 100. History

UniProtKB Accession: Q06609

Contribute: Send feedback

Clusters with 100%, 90%, 50% identity | Documents (... DF/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** Short name=hRAD51 Short name=HsRAD51 |
| Gene names | Name: **RAD51** Synonyms: RAD51A, RECA |
| Organism | **Homo sapiens (Human)** |
| Taxonomic identifier | 9606 [NCBI] |
| Taxonomic lineage | Eukaryota › Me... ni › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo |

Gene Name: RAD51

Synonyms: RAD51A

Taxon id: 9606 (human)

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.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

## 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 chos* *try refers to it. This is also the sequence that appears in the downloadable versions of the entry.*

Isoform identifiers

**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.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**<br>Secondary accession number(s): Q6FHX9, Q9BV60 |
| Entry history | Integrated into UniProtKB/Swiss-Prot: June 1, 1994<br>Last sequence update: June 1, 19<br>Last modified: Septembe<br>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) |

Secondary (retired) identifiers

## Relevant documents

Hide | Top

Human chromosome 15
Human chromosome 15: entries, gene names and cross-references to MIM

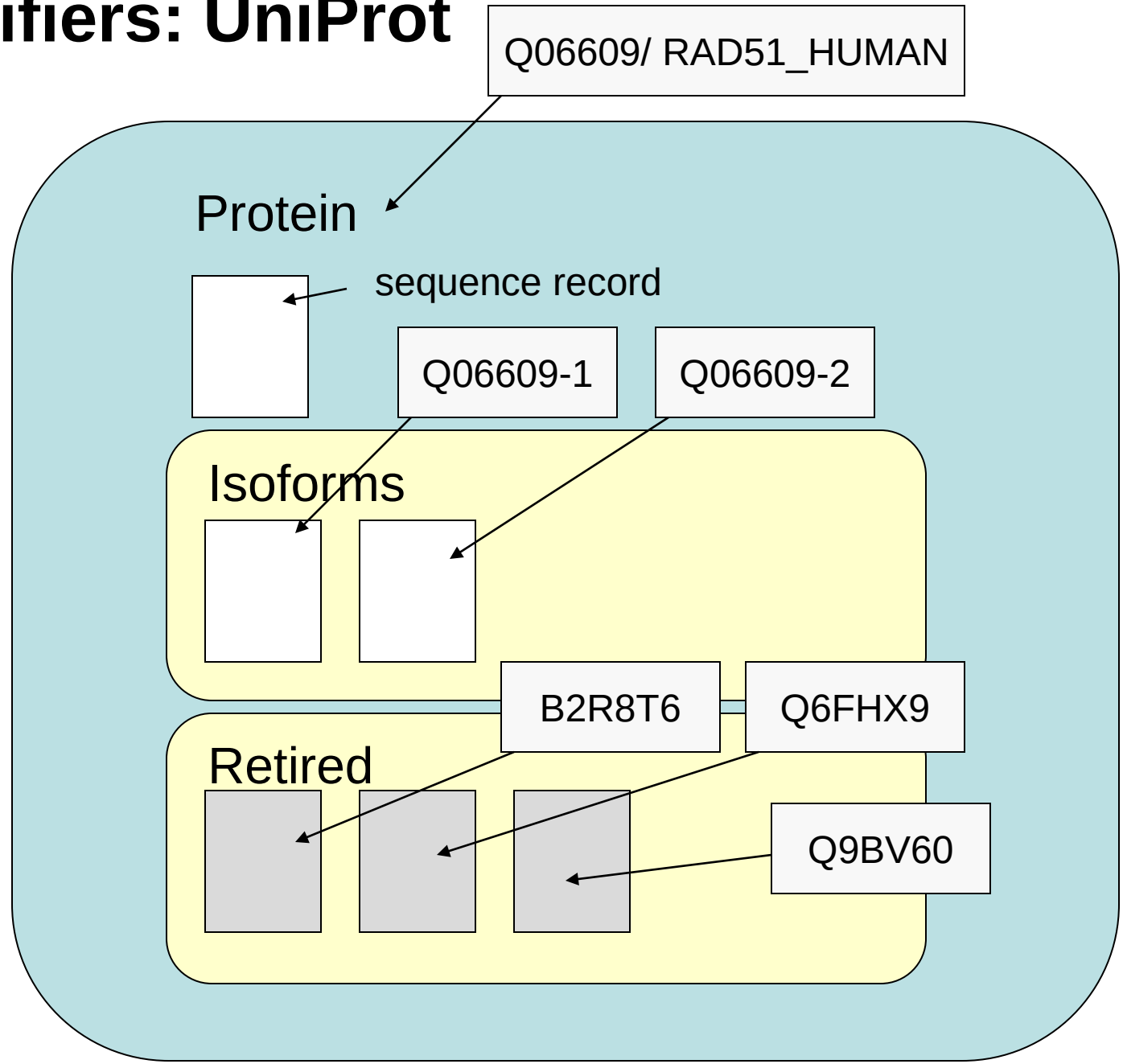
Human entries with polymorphisms or disease mutations
List of human entries with polymorphisms or disease mutations

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



For your information

# 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**

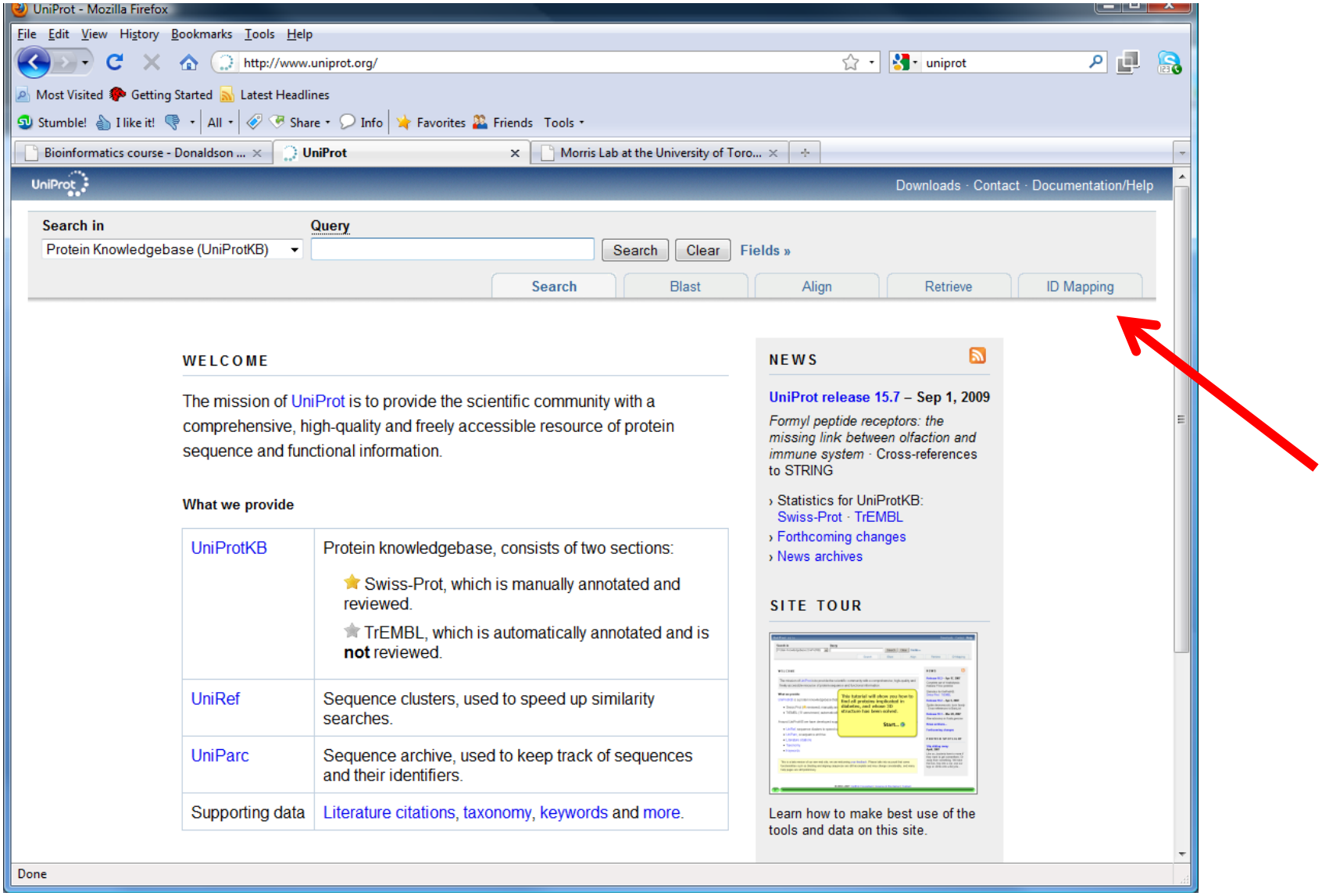
InterPro IPR015252
OMIM 600185
Pfam PF09104
Gene Ontology GO:0000724
SNPs rs28897757

**Experimental Platform**

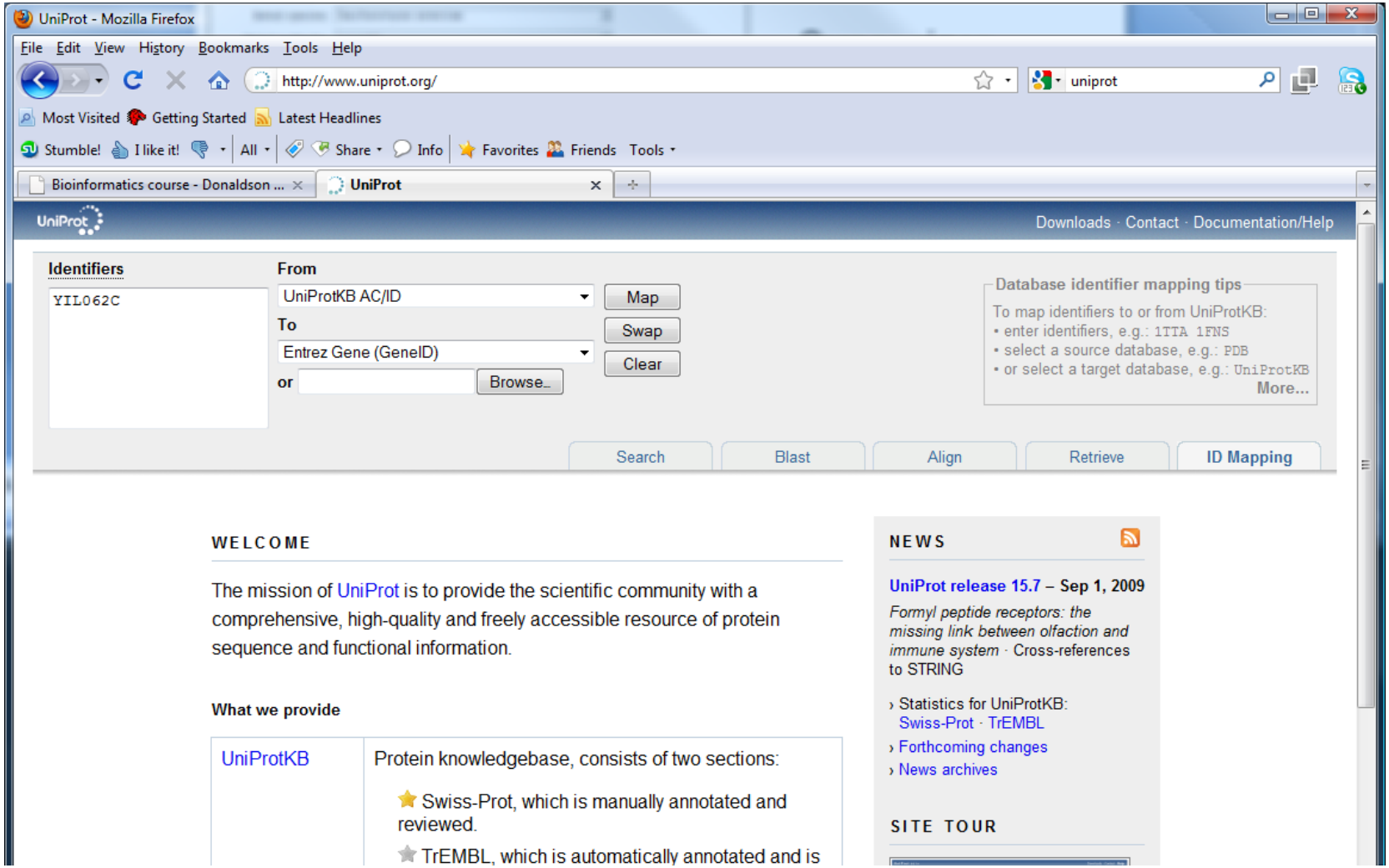
Affymetrix 208368_3p_s_at
Agilent A_23_P99452
CodeLink GE60169
Illumina GI_4502450-S

Red = Recommended

# UniProt ID Mapping Service

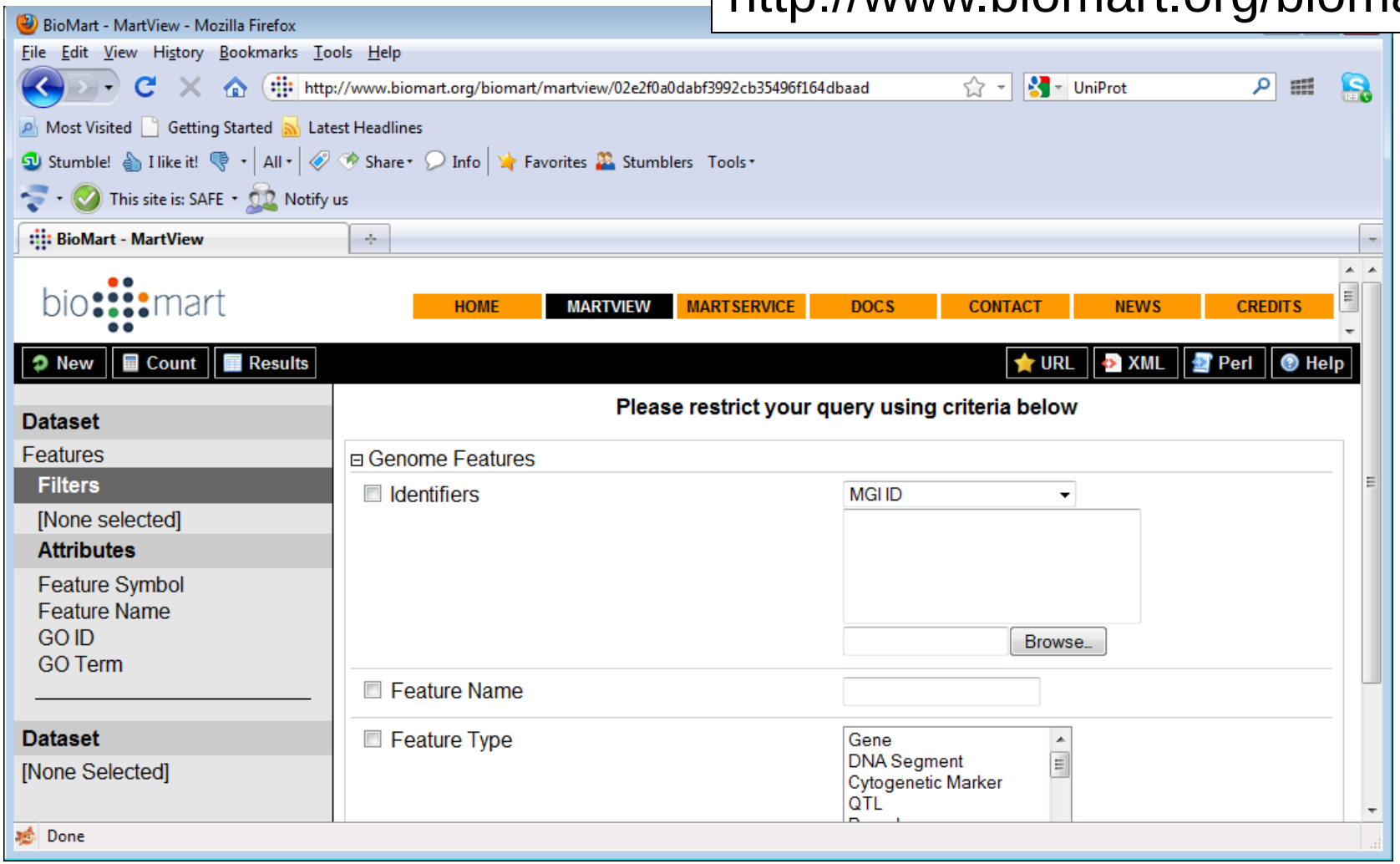


# UniProt ID Mapping Service



# Ensembl BioMart

http://www.biomart.org/biomart



# 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)**