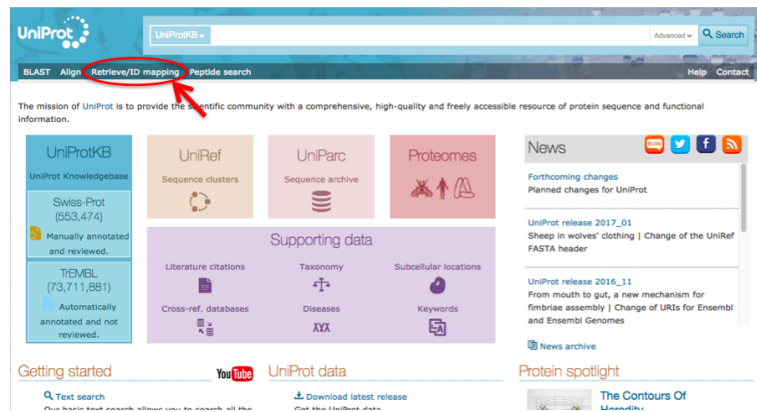


## Proteomics Exercise Week 6 Part 1

1. Go to [www.uniprot.org](http://www.uniprot.org) and select the "Retrieve/ID mapping" link.

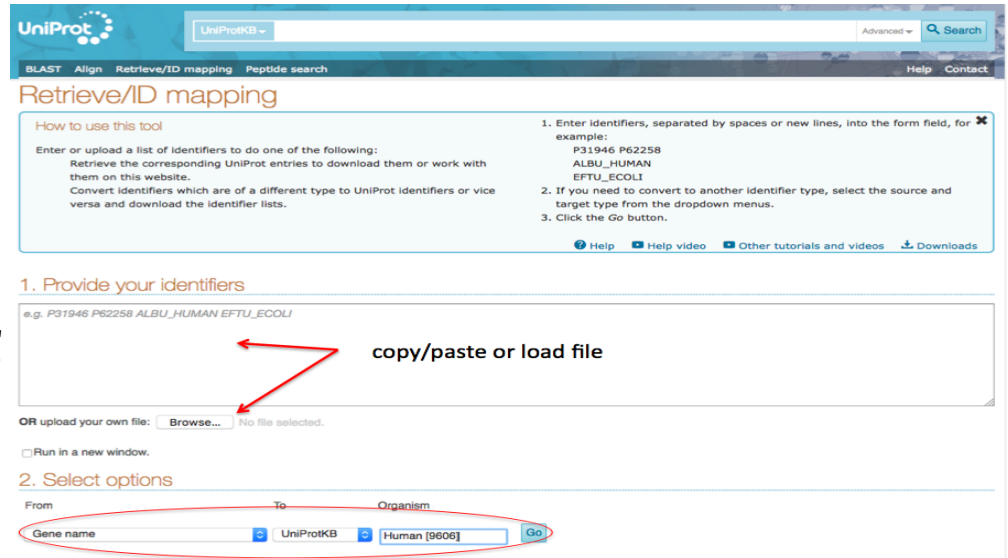


2. Open the file "ProteomicsDemoList1.txt". Copy/paste this list of identifiers into the box, or browse and load the file.

Select the options "Gene Name" to "UniProtKB" and Organism = "Human" [9606].

Tip: Be careful not to select a "Human Virus" or another organism with "Human" as part of the name.

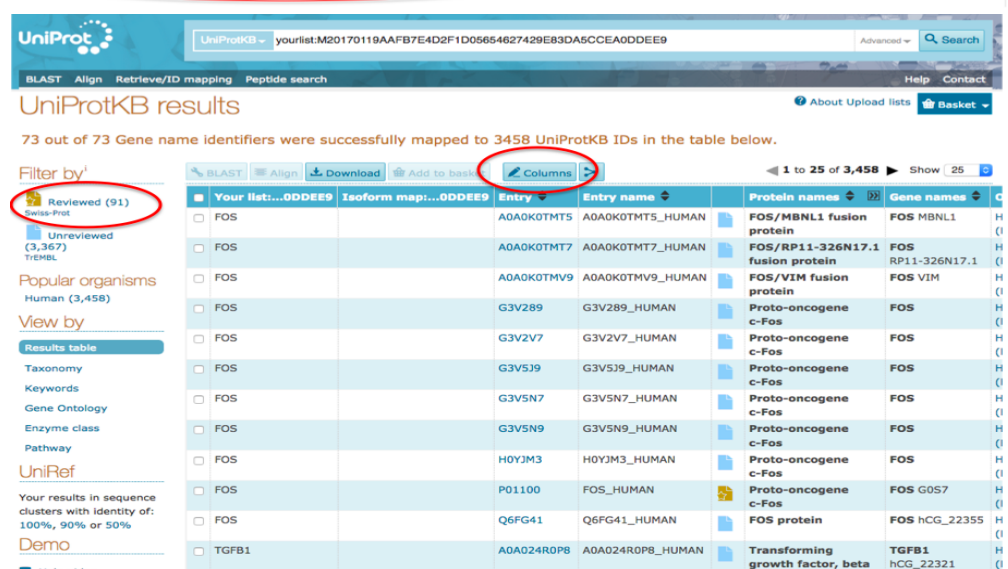
Then select "go."



3. Observe the results.

Select filter by "Reviewed" to limit to Human Reviewed Entries. The reasons for this step are covered in the Demo.

After filtering, select the "Columns" button.



4. Edit the columns as described in the demo and UniProt demo video on online help.

Remove the "isoform mapping" column.

Add the "Gene names (primary)" column.

Reorder the columns so that "Your list..." and "Gene names (primary)" are next to each

other and near the first columns.

Select "Save."

### Customize results table

Columns to be displayed<sup>1</sup> Reset to default Save Cancel

Drag and drop to re-order.

Enty Your list:...0DDEE9 Gene names (primary) Gene names Entry name Protein names Organism Length

Add more columns<sup>1</sup> Expand all

Search: e.g. gene, ontology,...

**Names & Taxonomy**

- ☒ Entry name
- ☒ Protein names
- ☒ Gene names
- ☒ Gene names (primary)
- ☐ Gene names (synonym)
- ☐ Gene names (ordered locus)
- ☐ Gene names (ORF)
- ☒ Organism ID
- ☐ Organism
- ☐ Proteomes
- ☐ Taxonomic lineage
- ☐ Virus hosts

**Sequences**

- ☒ Length
- ☐ Fragment
- ☐ Gene encoded by
- ☐ Alternative products (isoforms)
- ☐ Erroneous gene model prediction
- ☐ Mass spectrometry
- ☐ Polymorphism
- ☐ RNA editing
- ☐ Sequence caution
- ☐ Mass
- ☐ Sequence
- ☐ Alternative sequence
- ☐ Natural variant
- ☐ Non-adjacent residues
- ☐ Non-standard residue

**Function**

- ☐ EC number
- ☐ Absorption
- ☐ Catalytic activity
- ☐ Cofactor
- ☐ Enzyme regulation
- ☐ Function [CC]<sup>1</sup>
- ☐ Kinetics
- ☐ Pathway
- ☐ Redox potential
- ☐ Temperature dependence
- ☐ pH dependence
- ☐ Active site
- ☐ Binding site
- ☐ Calcium binding
- ☐ DNA binding
- ☐ Metal binding

**Miscellaneous**

- ☐ Annotation score
- ☐ Features
- ☐ Caution
- ☐ Miscellaneous [CC]<sup>1</sup>
- ☐ Isoform map:...0DDEE9
- ☐ Isoform map:...CB7771
- ☐ Isoform map:...BD1876
- ☐ Keywords
- ☐ Matched text
- ☐ PeptideSearch:...1FB3CX
- ☐ Protein existence
- ☐ Tools
- ☐ UniParc
- ☒ Your list:...0DDEE9

5. Select all entries.

We will now review the list for incorrect matches to old gene names or duplicate matches.

We will deselect the entries we do not want. See the demo video for examples.

When done, select "Add to basket." Then select "Basket," then "Full View."

You should have 73 IDs mapped to 73 UniProt Entries.

UniProtKB results

73 out of 73 Gene name identifiers were successfully mapped to 91 UniProt Entries.

Filter by<sup>1</sup> Align Download Add to basket Columns

Reviewed (91) Entry Your list:...0DDEE9 Gene names (primary)

Popular organisms Human (91)

View by Results table Taxonomy Keywords Gene Ontology Enzyme class Pathway UniRef

Your results in sequence clusters with identity of: 100%, 90% or 50%

Entry	Gene names (primary)	Gene names	Entry name	Protein names	Length	Function [CC]
P01100	FOS	FOS	FOS_HUMAN	Tumor necrosis factor	233	NP_001
P01137	TGFB1	TGFB1	TGFB1_HUMAN	Interferon alpha-2	188	NP_001
P01375	TNF	TNF	TNFA_HUMAN	Interleukin-1 beta	269	NP_001
P01563	IFNA2	IFNA2	IFNA2_HUMAN	Alpha-crystallin B chain	175	NP_001
P01584	IL1B	IL1B	IL1B_HUMAN			
P02511	CRYAB	CRYAB	CRYAB_HUMAN			

UniProtKB (73) UniRef (91) UniParc (91)

Download selected (0) Download all (73)

Format: Excel FASTA (canonical) FASTA (canonical & isoform) Tab-separated Text Preview

1 to 73 of 73 Show 100

Gene names	Entry name	Protein names	Length	Func
YWHAQ	1433T_HUMAN	14-3-3 protein theta	245	Adapti

### Customize results table

Columns to be displayed<sup>1</sup> Reset to default Save Cancel

Drag and drop to re-order.

Entry Your list:...0DDEE9 Gene names (primary) Gene names Entry name Protein names Length Function [CC]

Involvement in disease Interacts with Gene ontology (biological process) Gene ontology (molecular function) Gene ontology (cellular component) HGNC

PubMed ID RefSeq Reactome Ensembl GeneID

Add more columns<sup>1</sup> Expand all

Search: e.g. gene, ontology,...

**Names & Taxonomy**

- ☒ Gene names
- ☒ Gene names (primary)
- ☒ Gene names (synonym)
- ☒ Gene names (ordered locus)
- ☒ Gene names (ORF)

**Sequences**

- ☒ Length
- ☐ Fragment
- ☐ Gene encoded by
- ☐ Alternative products (isoforms)
- ☐ Erroneous gene model prediction

**Function**

- ☐ EC number
- ☐ Absorption
- ☐ Catalytic activity
- ☐ Cofactor
- ☐ Enzyme regulation
- ☐ Kinetics

**Miscellaneous**

- ☐ Annotation score
- ☐ Features
- ☐ Caution
- ☐ Miscellaneous [CC]<sup>1</sup>

7. Select "Download."

Select "uncompressed."

Select "Excel or tab-separated."

That is it. You have now mapped your identifiers to other resources and added functional information.

Save this Excel sheet—we will use this information in the next part of this exercise.

UniProtKB results

73 out of 73 UniProtKB AC/ID identifiers were successfully mapped to 73 UniProtKB IDs in the table below.

Filter by<sup>1</sup> Align Download Add to basket Columns

Reviewed (73) Entry Your list:...0DDEE9 Gene names (primary)

Popular organisms Human (73)

View by Results table Taxonomy Keywords Gene Ontology Enzyme class Pathway UniRef

Your results in sequence clusters with identity of: 100%, 90% or 50%

Entry	Gene names (primary)	Gene names	Entry name	Protein names	Length	Func
P27348	YWHAQ	YWHAQ	1433T_HUMAN	14-3-3 protein theta	245	Adapti

Download selected (0) Download all (73)

Format: Excel FASTA (canonical) FASTA (canonical & isoform) Tab-separated Text Preview

1 to 73 of 73 Show 100

Gene names	Entry name	Protein names	Length	Func
YWHAQ	1433T_HUMAN	14-3-3 protein theta	245	Adapti