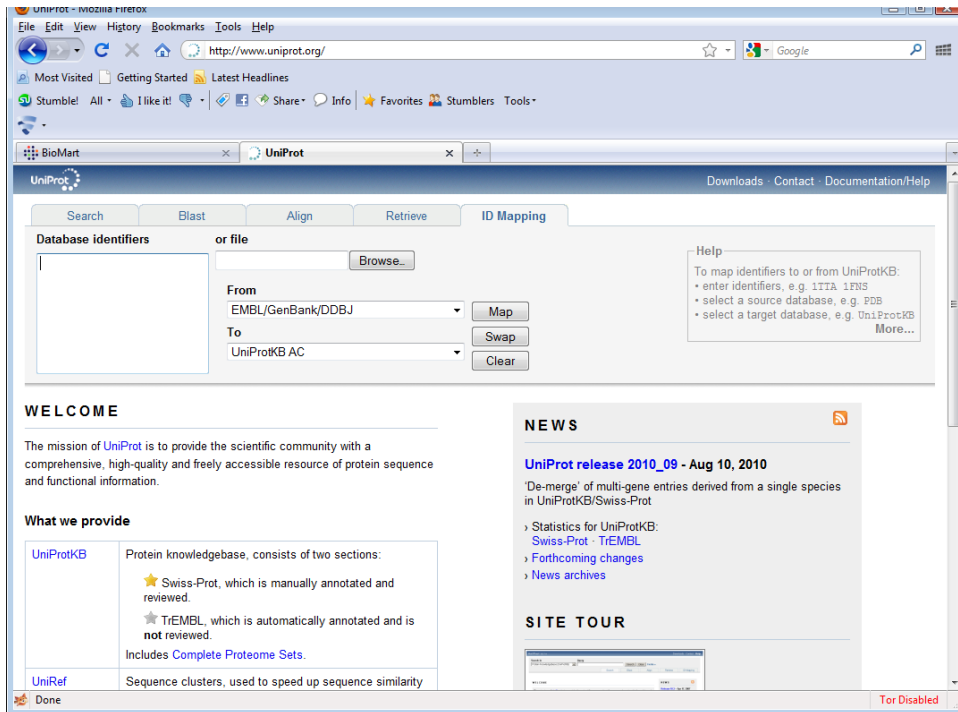


## Identifier conversion exercise

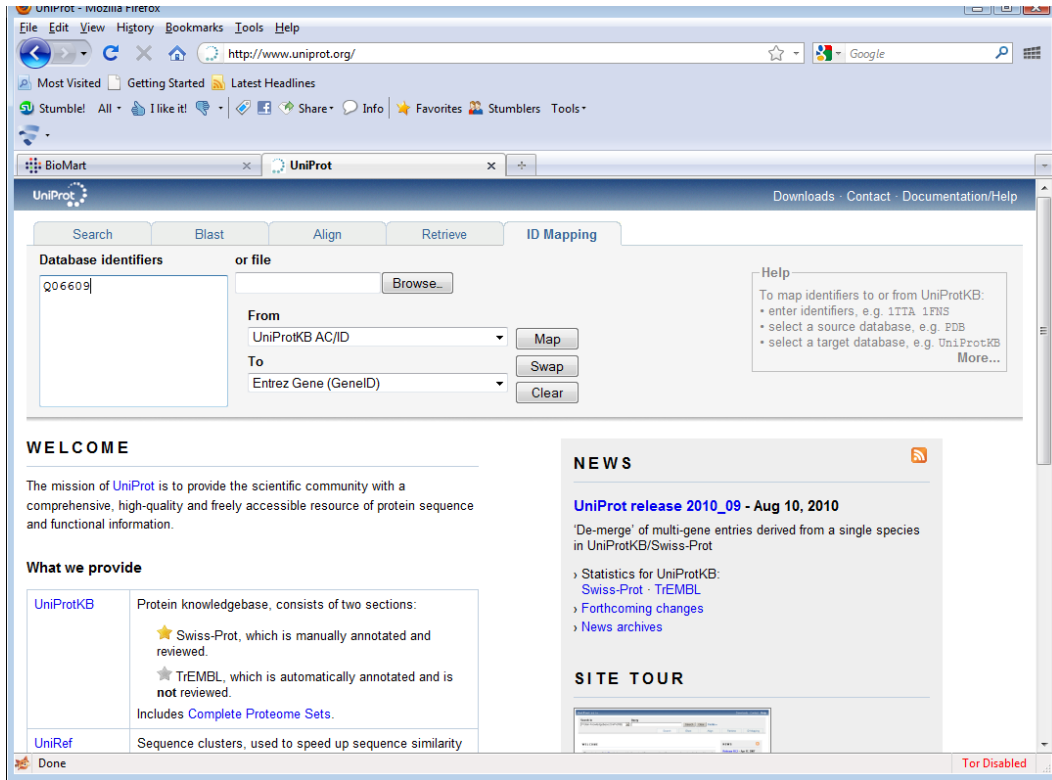
MBV-INF4410/9410

## UniProt ID Mapping service

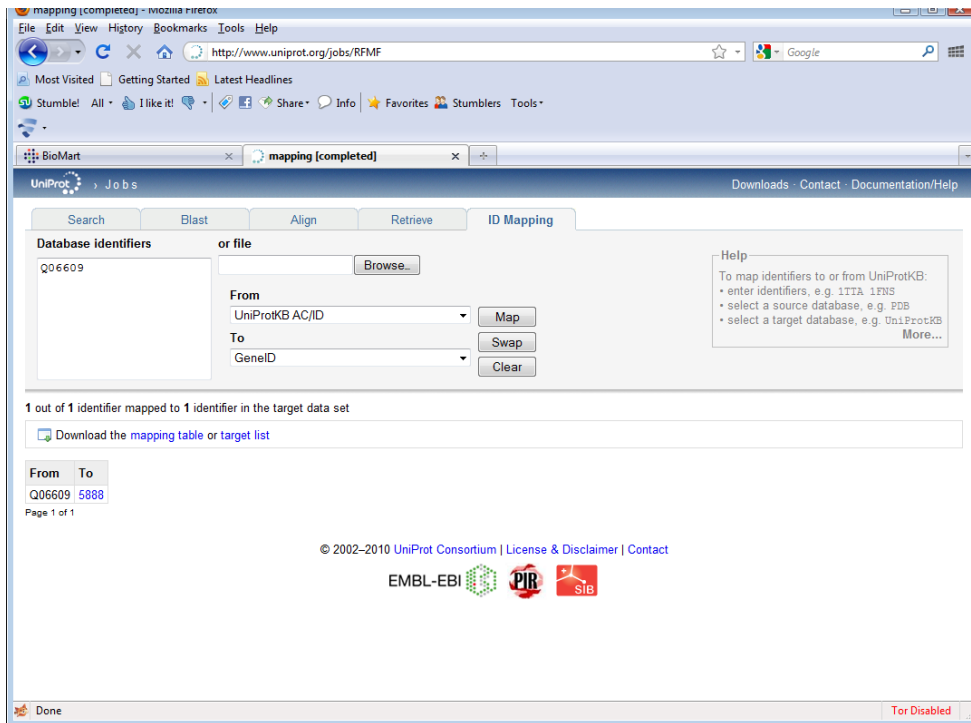
1. Go to <http://www.uniprot.org/>
2. Click on ID Mapping



3. Enter "Q06609" under database identifiers and select From: UniProt AC/ID and TO: Entrez Gene ID. Like this.



4. Click on “Map” and you should see this:

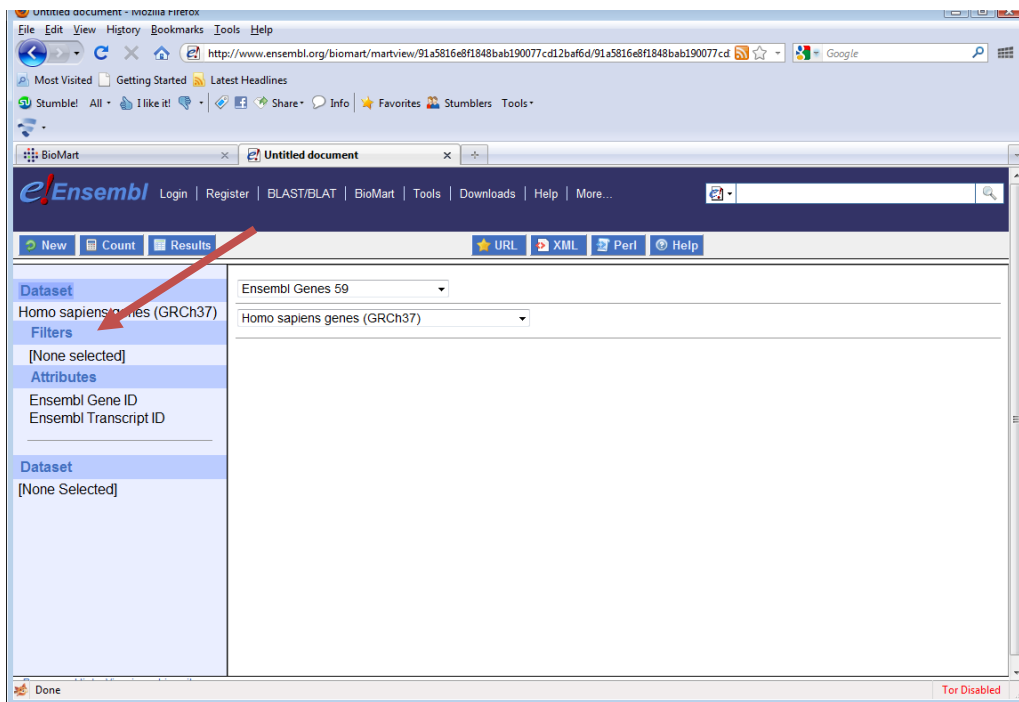


5. You can export results to mapping table for import into Excel or some other application.

- Obviously, this can be down for a list of identifiers that need to be converted.
- Try converting to other identifier types and validate your results by going to the source database. Does Entre Gene record 5888 refer to the same thing as your query? Does this record make a cross-reference back to UniProt record **Q06609**?

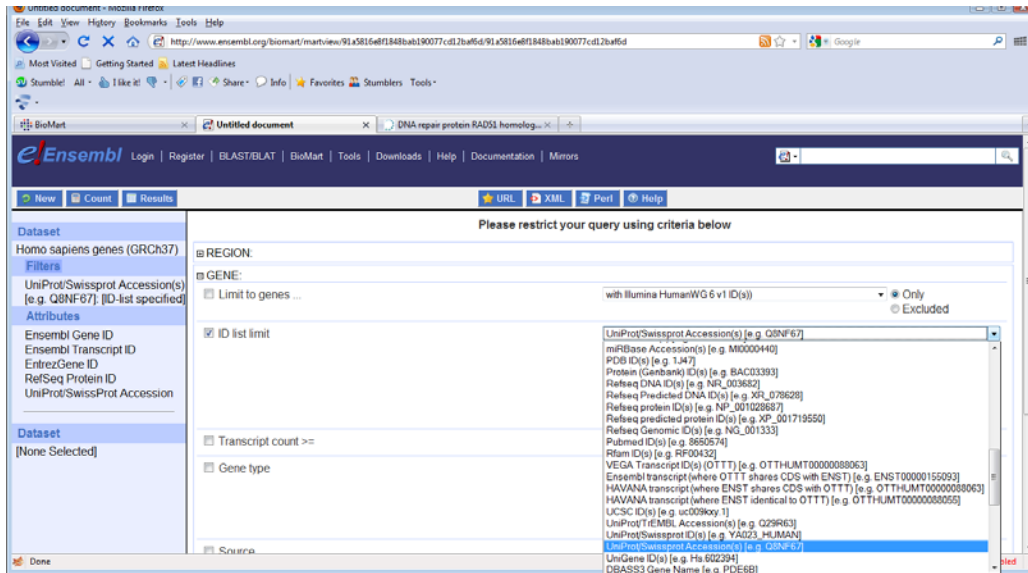
## BioMart

- Go to <http://www.biomart.org/>
- Click on Ensembl to go to <http://www.ensembl.org/biomart/martview/>
- Under “choose database”, select ENSEMBL Genes 59
- Under “choose dataset”, select “Homo sapiens genes (GRCh37)”

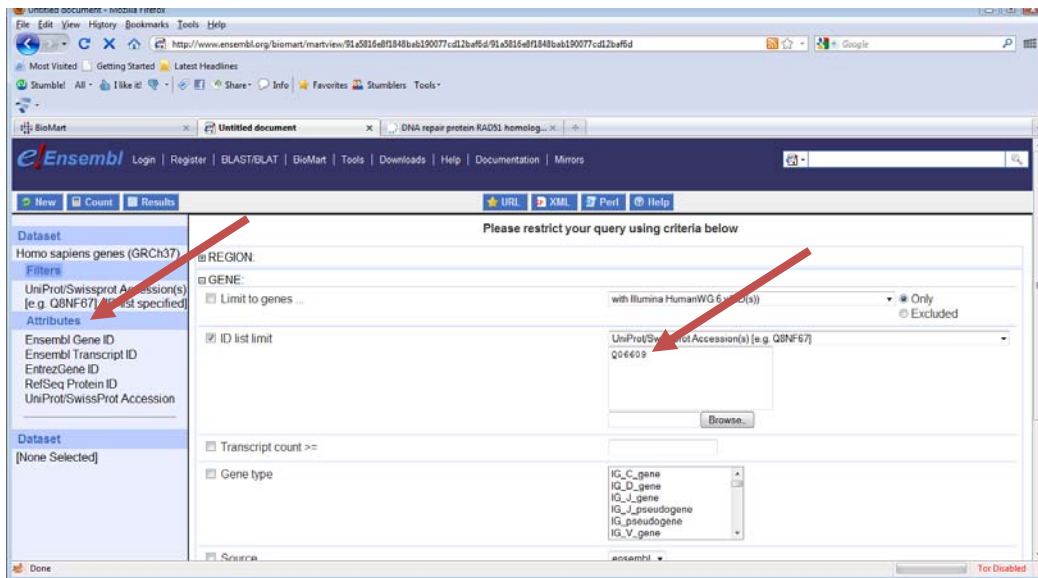


- Click on filters (this is where you will specify what you are looking for)
- Click on + sign beside “Gene”

7. Check the box beside ID list limit and select UniProtAccession(s) from the drop down list

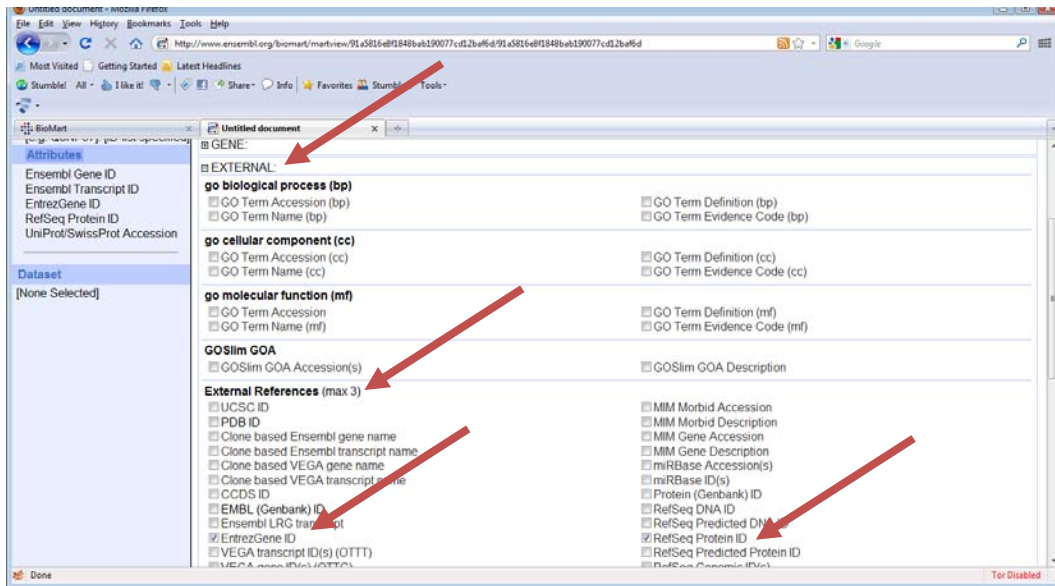


8. Enter "Q06609" in the box

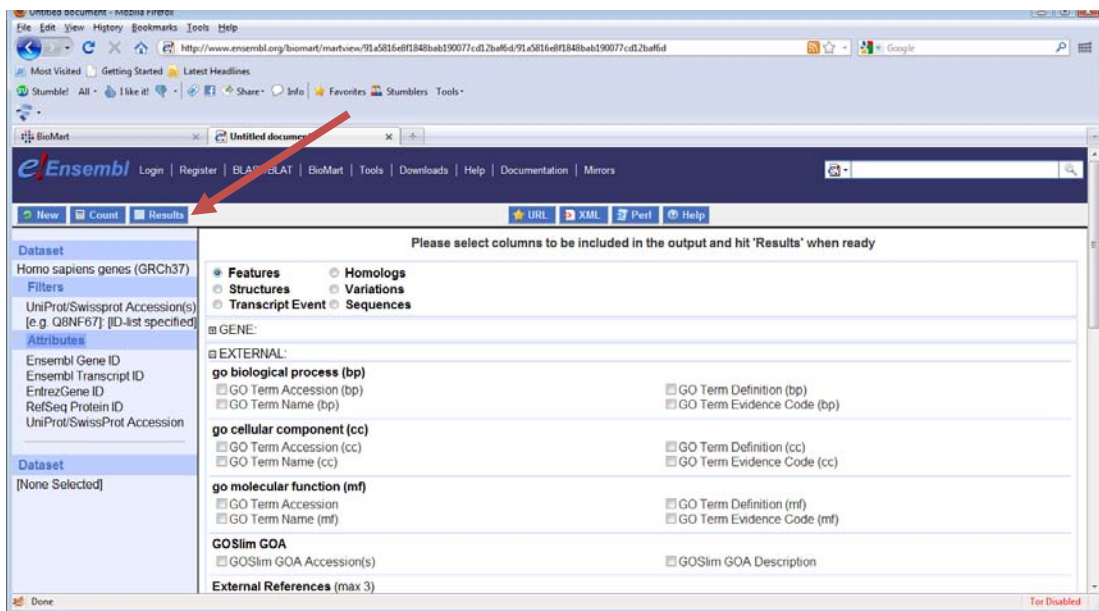


9. Click on attributes. This is where you will specify what information you want from the records that your search finds.

10. Then on external and then (under External references) select  
Entrez Gene ID  
RefSeq Protein ID and  
UniProt/SwissProt Accession



11. Click on results in the upper left hand corner



12. You should see results like these

The screenshot shows the Ensembl BioMart interface. The browser address bar displays the URL: <http://www.ensembl.org/biomart/martview/9145816ef1848bab19077cd12ba6fd9145816ef1848bab19077cd12ba6fd>. The interface includes a navigation bar with links for Login, Register, BLAST/ELAT, BioMart, Tools, Downloads, Help, Documentation, and Mirrors. Below this, there are buttons for New, Count, Results, and a toolbar with options for URL, XML, Perl, and HTML. The main content area is divided into a left sidebar and a main panel. The sidebar contains sections for Dataset (Homo sapiens genes (GRCh37)), Filters (UniProt/Swissprot Accession(s) [e.g. Q8NF67] [ID-list specified]), Attributes (Ensembl Gene ID, Ensembl Transcript ID, EntrezGene ID, RefSeq Protein ID, UniProt/SwissProt Accession), and Dataset ([None Selected]). The main panel features an 'Export all results to' section with a dropdown menu set to 'File', a 'TSV' radio button selected, and a 'Unique results only' checkbox. Below this is an 'Email notification to' field. The 'View' section shows '10 rows as HTML' and a 'Unique results only' checkbox. A table displays the following data:

Ensembl Gene ID	Ensembl Transcript ID	EntrezGene ID	RefSeq Protein ID	UniProt/SwissProt Accession
<a href="#">E43Q200000001100</a>	<a href="#">E43Q21000000007000</a>	<a href="#">2889</a>	<a href="#">NP_002089</a>	<a href="#">Q05609</a>

Results can be exported as tab-separated lists (TSV), as an excel file (XLS).

Try retrieving other information about this same sequence record (like GO terms).