## Identifier conversion exercise

MBV-INF4410/9410

UniProt ID Mapping service

- 1. Go to <a href="http://www.uniprot.org/">http://www.uniprot.org/</a>
- 2. Click on ID Mapping

UNIPROT - MOZIIIA FIREFOX				
<u>Eile Edit V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp				
C X A A http://www.uniprot.org/		☆ 🔹 🚰 🕇 Google	م	==
🙍 Most Visited 📄 Getting Started 🔜 Latest Headlines				
💿 Stumble! All - 🍐 I like it! 👎 - 🛷 🖪 🧇 Share - 🔎 Info 🎽 Favorites 🗸	Stumblers Tools*			
17 - C				
BioMart × UniProt	× +			-
UniProt		Downloads · Contact · Documer	ntation/Help	-
Search Blast Align Retrieve	ID Mapping			
Database identifiers or file				
Browse_		Help		
From		To map identifiers to or from UniPr • enter identifiers, e.g. 1TTA 1FNS	OTKB:	
EMBL/GenBank/DDBJ	<ul> <li>Map</li> </ul>	<ul> <li>select a source database, e.g. P</li> <li>select a target database, e.g. Un</li> </ul>		E
То		<ul> <li>select a target database, e.g. un</li> </ul>	More	
UniProtKB AC	Swap			
	Clear			
				1
WELCOME	NEWS	<b>N</b>		
The mission of UniProt is to provide the scientific community with a				
comprehensive, high-quality and freely accessible resource of protein sequence	UniProt release 20	010_09 - Aug 10, 2010		
and functional information.		ne entries derived from a single species		
	in UniProtKB/Swiss-P	rot		
What we provide	<ul> <li>Statistics for UniProt</li> </ul>			
UniProtKB Protein knowledgebase, consists of two sections:	Swiss-Prot · TrEMBL > Forthcoming changes			
Swiss-Prot, which is manually annotated and	> News archives			
reviewed.				
TrEMBL, which is automatically annotated and is not reviewed.	SITE TOUR			
Includes Complete Proteome Sets.		tana ara ara		
UniRef Sequence clusters, used to speed up sequence similarity	Hann Bar Wil (MM	An New Descent		-
😸 Done			Tor Disable	ed _

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

UNIProt - Iviozilia Firetox			<u> </u>
<u>File Edit View History Bookmarks T</u> ools <u>H</u> elp			
C X A D http://www.uniprot.org/		Google	=
🙇 Most Visited 📄 Getting Started 🔜 Latest Headlines			
💿 Stumble! All + 🏠 I like it! 👎 + 🛷 🖪 🧇 Share+ 💭 Info 🎽 Favorites 🖁	Stumblers Tools+		
1971 - C.			
BioMart × UniProt	x *		-
UniPro		Downloads · Contact · Documentation/Help	
Search Blast Align Retrieve	ID Mapping		
Database identifiers or file		U-la	
Q06609 Browse_		- Help To map identifiers to or from UniProtKB:	
From		<ul> <li>enter identifiers, e.g. 1TTA 1FNS</li> </ul>	
UniProtKB AC/ID		<ul> <li>select a source database, e.g. PDB</li> <li>select a target database, e.g. UniProtKB</li> </ul>	Ξ
Te	Swap	More	
Entrez Gene (GeneID)			
	Clear		
WELCOME	NEWS		
The mission of UniProt is to provide the scientific community with a			
comprehensive, high-quality and freely accessible resource of protein sequence	UniProt release 2010_09 - /	Aug 10. 2010	
and functional information.	'De-merge' of multi-gene entries	-	
	in UniProtKB/Swiss-Prot	5 .	
What we provide	> Statistics for UniProtKB:		
UniProtKB Protein knowledgebase, consists of two sections:	Swiss-Prot · TrEMBL		
	> News archives		
Swiss-Prot, which is manually annotated and reviewed.			
TrEMBL, which is automatically annotated and is not reviewed.	SITE TOUR		
Includes Complete Proteome Sets.	2000/000/000/000/000/000/000/000/000/00	-	
UniRef Sequence clusters, used to speed up sequence similarity	100 00 00 000 000	80. 8	-
😹 Done	A A A A A A A A A A A A A A A A A A A	Tor Disable	d

4. Click on "Map" and you should see this:

_ 🤓 mapping [completeα] - Ινιοzilia Firetox	
Eile Edit View Higtory Bookmarks Iools Help	
C X 🟠 💭 http://www.uniprot.org/jobs/RFMF	Google
🙍 Most Visited 📋 Getting Started 🔜 Latest Headlines	
💿 Stumble! All + 🍐 I like it! 🤻 + 🖉 🖪 🥐 Share+ 🔎 Info 🎽 Favorites 🚨 Stumblers 🛛 Tools+	
BioMart × mapping (completed) × ÷	
UniProt Jobs Download	ds · Contact · Documentation/Help
Search Blast Align Retrieve ID Mapping	
Database identifiers or file	
From • enter iden	ntifiers to or from UniProtKB: tifiers, e.g. ITTA IFNS source database, e.g. DDB arget database, e.g. UniProtKB More
1 out of 1 identifier mapped to 1 identifier in the target data set	
From To Q06609 5888 Page 1 of 1	
© 2002–2010 UniProt Consortium   License & Disclaimer   Contact EMBL-EBI	
👏 Done	Tor Disabled

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

- 6. Obviously, this can be down for a list of identifiers that need to be converted.
- 7. 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

- 1. Go to <a href="http://www.biomart.org/">http://www.biomart.org/</a>
- 2. Click on Ensemble to go to <a href="http://www.ensembl.org/biomart/martview/">http://www.ensembl.org/biomart/martview/</a>
- 3. Under "choose database", select ENSEMBL Genes 59
- 4. Under "choose dataset", select "Homo sapiens genes (GRCh37)"

Untitled document - Mozilia Firefox		
<u>File Edit View History Bookmarks Too</u>		
	r://www.ensembl.org/biomart/martview/91a5816e8f1848bab190077cd12baf6d/91a5816e8f1848bab190077cd 🗟 😭 🕤 🛃 🐔 Google	= م
🔊 Most Visited 🗋 Getting Started 🔊 Late	est Headlines	
	🖪 🎯 Share+ 🔎 Info 🙀 Favorites 🎎 Stumblers 🛛 Tools+	
7		
BioMart ×	C P Untitled document X ÷	-
		·
CEnsembl Login   Regi	ister   BLAST/BLAT   BioMart   Tools   Downloads   Help   More 🛃 -	9
-		
New Count Results	🛨 URL 🛛 XML 🔯 Perl 🔞 Help	
Dataset	Ensembl Genes 59 -	
Homo sapienstornes (GRCh37)	Homo sapiens genes (GRCh37)	
Filters	Homo sapiens genes (GRCh37)	
[None selected]		
Attributes		
Ensembl Gene ID		
Ensembl Transcript ID		=
Dataset		
[None Selected]		
		-
👏 Done		Tor Disabled

- 5. Click on filters (this is where you will specify what you are looking for)
- 6. Click on + sign beside "Gene"

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

Untituea document - Mozilia Firefox			
Eile Edit View Higtory Bookmarks Io			
🕙 🕗 C 🗙 🏠 🖻 🔤	r://www.ensembl.org/biomart/martview/91a5816e8f1848bab190077cd12ba/6d/91a5816e8f1848bab19007	/cd12baf6d 🔂 🏫 🔹 🚼 🔹	Google 🔎 🖩
🔬 Most Visited 📄 Getting Started <u>a</u> Lat	est Headlines		
型 Stumble! All - 🍐 I like it! 🤻 - 🏼 🌜	🛙 🖪 🕐 Share - 💭 Info 🙀 Favorites 🎇 Stumblers - Tools -		
7			
-	C Untitled document X DNA repair protein RAD51 homolog X +		
ngi bromat			
CEnsembl Login   Rog	jister   BLAST/BLAT   BioMart   Tools   Downloads   Help   Documentation   Mirrors	📇 -	9
🤉 New 🔋 Count 🔳 Results	🝁 URL 🜗 XML 🚦	Peri 🔿 Help	
Dataset	Please restrict your	query using criteria below	
Homo sapiens genes (GRCh37)	B REGION:		
UniProt/Swissprot Accession(s)	ID GENE:		
[e.g. Q8NF67]: [ID-list specified]		with Illumina HumanWG 6 v1 ID(s))	<ul> <li>Only</li> </ul>
Attributes			Excluded
Ensembl Gene ID	ID list limit	UniProt/Swissprot Accession(s) [e.g. Q8NF67]	-
Ensembl Transcript ID		miRBase Accession(s) [e.g. MI0000440]	•
EntrezGene ID		PDB ID(s) [e.g. 1J47] Protein (Genbank) ID(s) [e.g. BAC03393]	
RefSeq Protein ID UniProt/SwissProt Accession		Refseq DNA ID(s) [e.g. NR_003682] Refseq Predicted DNA ID(s) [e.g. XR_078628]	
UNPTO SWISSPICE ACCESSION		Refseg protein ID(s) [e.g. NP_001028687]	
		Refseq predicted protein ID(s) [e.g. XP_001719550] Refseq Genomic ID(s) [e.g. NG_001333]	
Dataset	Transcript count >=	Pubmed ID(s) [e.g. 8650574]	
[None Selected]	E Outstaa	Rfam ID(s) [e.g. RF00432] VEGA Transcript ID(s) (OTTT) [e.g. OTTHUMT00000	0890631
	Gene type	Ensembl transcript (where OTTT shares CDS with EN	ST) [e.g. ENST00000155093]
		HAVANA transcript (where ENST shares CDS with OT HAVANA transcript (where ENST identical to OTTT) [	
		UCSC ID(s) [e.g. uc009kxy.1]	g of the second s
		UniProt/TrEMBL Accession(s) [e.g. Q29R63] UniProt/Swissprot ID(s) [e.g. YA023_HUMAN]	
	E Source	UniProt/SwissprotAccession(s) [e.g. Q8NF67]	
🖆 Done		UniGene ID(s) [e.g. Hs.602394] DBASS3 Gene Name (e.g. PDE6B)	pled

8. Enter "**Q06609**" in the box

Be Edit View History Bookmarks Too		e8f1848bab190077cd12baf5d/91a5816e8f1848bab190077cd12baf5d	🔂 🗘 - 🛃 - Geogle	P #
		811848bab190077cd12ba1bd/91a38104811848bab190077cd12ba1bd	E LI *	· // II
Most Visited 🧾 Getting Started 🔜 Later				
	🛐 🧐 Share - 💭 Info 🎽 Favorites 🎎 Stu	mblers Tools+		
Ş.				
i뷰: BioMart ×	C Untitled document ×	💭 DNA repair protein RADS1 homolog 🛪 🛛 🔶		
	ter   BLAST/BLAT   BioMart   Tools	Downloads   Help   Documentation   Mirrors	- -	a
			Charles de	aresta.
🤊 New 🔲 Count 🔲 Results		🝁 URL 😰 XML 😰 Perl 🕐 Help		
Dataset		Please restrict your query using crite	ria below	
Iomo sapiens genes (GRCh37) Filters	REGION:			
UniProt/Swissprot Accession(s)	© GENE:			
[e.g. Q8NF67] [Parst specified] Attributes	Limit to genes	with Ilumina Hum	anWG 6 yr 2(s))	Only     Excluded
Ensembl Gene ID	🗵 ID list limit	UniProt/Sweetro	Accession(s) [e.g. QSNF67]	•
Ensembl Transcript ID EntrezGene ID RefSeq Protein ID UniProt/SwissProt Accession		206609 <b>-</b>	Browse.	
Dataset None Selected]	Transcript count >=			
none selecieuj	Gene type	IG, C, gene IG, D, gene IG, J, gene IG, J, pseudogene IG, pseudogene IG, V, gene	0	
	E Source	ensembl •	12	
Done				Tor Disabled

- 9. Click on attributes. This is where you will specify what information you want from the records that your search finds.
- Then on external and then (under External references) select Entrez Gene ID RefSeq Protein ID and UniProt/SwissProt Accession

Contribed document - Mozilia Piletox			0.0	1.10
Eile Edit View Higtory Bookmarks Ic	ols Help			
🔇 🕞 C 🗙 🏠 🖻 🖛		d 🐻 😭 🔹 🚮 🖷 Google	P	Ħ
🖉 Most Visited 📄 Getting Started 📥 Lat	est Headlines			
🔮 Stumblel All - 🍐 Hike itl 🧐 - 🤞	💼 🥙 Share+ 💭 Info 🎽 Favorites 🚨 Stumble 🛛 Tools+			
4.				
the BioMart	😰 Untitled document 🛛 🗴 🔅			-
for 3. more out for not obcounted	B GENE:			
Attributes	# EXTERNAL:			
Ensembl Gene ID	go biological process (bp)			
Ensembl Transcript ID EntrezGene ID	G G Term Accession (bp)	GO Term Definition (bp)		
RefSeq Protein ID UniProt/SwissProt Accession		El GO Term Evidence Code (op)		
	go cellular component (cc) CO Term Accession (cc)	GO Term Definition (cc)		
Dataset	I GO Term Name (cc)	GO Term Evidence Code (cc)		
[None Selected]	go molecular function (mf) CO Term Accession GO Term Name (mf)	CO Term Definition (mf)		
	GOSIim GOA GOSIim GOA Accession(s)	COSIIm GOA Description		
	External References (max 3) UCSC ID UCSC ID Cone based Ensembl gene name Cone based Ensembl transcript name Cone based VEGA gene name Cone based VEGA gene name Cone based VEGA transcript name Cone Based transcript name	MIM Morbid Accession MIM Gene Accession MIM Gene Description MIM Gene Description mrRBase Accession(s) mrRBase D(s) Protein (Genbank) ID RetSeq DNA ID RetSeq Prodicted DNA RetSeq Predicted DNA RetSeq Predicted Protein ID		
😸 Done	IIVECA association	III DefCas Casamia ID/s)	Tor Disable	d

11. Click on results in the upper left hand corner

Ble Edit View History Bookmarks To	and in Made		
	oonpp p://www.ensembl.org/biomart/martview/91a5816e8f1848bab190077cd12baf6d/91a5816e8f1848bab190077cd1	2baffd 🔯 🚽 🛃 * Google	P
Most Visited Getting Started			
	est reactions Info   🙀 Favorites 🌇 Stumblers Tools-		
5			
Statistics and the second s	X C Untitled document X		
	gister   BLACTIBLAT   BioMart   Tools   Downloads   Help   Documentation   Mirrors	8.	9
			10.70
🤊 New 🗧 Count 📓 Results 🖌	🔶 URL 📴 XML 🛐 P	erl 🔞 Help	
Dataset	Please select columns to be included in	n the output and hit 'Results' when ready	
Homo sapiens genes (GRCh37) Filters UniProt/Swissprot Accession(s)	Features     Homologs     Structures     Variations     Transcript Event     Sequences		
[e.g. Q8NF67]: [ID-list specified]			
Attributes	REXTERNAL:		
Ensembl Gene ID Ensembl Transcript ID	go biological process (bp)		
EntrezGene ID RefSeg Protein ID	GO Term Accession (bp)	GO Term Definition (bp) GO Term Evidence Code (bp)	
UniProt/SwissProt Accession	go cellular component (cc)		
Dataset	GO Term Accession (cc)	GO Term Definition (cc)	
None Selected]	go molecular function (mf)		
	GO Term Accession	GO Term Definition (mf)     GO Term Evidence Code (mf)	
	COSIm GOA	GOSIm GOA Description	
	External References (max 3)		
Done			Tor Disabled

12. You should see results like these

Unbitied document - Mozilia Firefox				
Eile Edit View Higtory Bookmarks To-				
🔇 🔄 🗣 😋 🕻 🕅 🗠	r://www.ensembl.org/biomart/martview/91a581	6e8f1848bab190077cd12baf6d/91a5816e8f1848bab190077cd12baf6d	📓 合 🔹 🚺 🖬 Google	##
🔊 Most Visited 🗋 Getting Started 🗟 Late	est Headlines			
型 Stumble! All - 🍐 I like it! 🤻 - 🔗	🗄 💽 🥗 Share - 💭 Info 🙀 Favorites 🏔 S	tumblers Tools-		
t∰: BioMart ×	Untitled document	DNA repair protein RAD51 homolog× +		-
CEnsembl Login   Rogi	ister   BLAST/BLAT   BioMart   Tools	Downloads   Help   Documentation   Mirrors	₫•	<u>م</u>
🔉 New 📓 Count 🔳 Results		🚖 URL 🔁 XML 🛃 Perl 🕅	) H/to	
Dataset	Export all results to	File • TSV •	Unique results only So	
Homo sapiens genes (GRCh37) Filters	Email notification to			
UniProt/Swissprot Accession(s)		10 👻 rows as HTML 👻 🖾 Unique results on	ly .	
[e.g. Q8NF67]: [ID-list specified] Attributes	Ensembl Gene ID Ensembl Transcr	pt ID EntrezGene ID RefSeq Protein ID UniProt/SwissPr	ot Accession	
Ensembl Gene ID	ENS00000051180 ENST0000026786	5888 NP_002866 Q06609		
Ensembl Transcript ID				1
EntrezGene ID				
RefSeq Protein ID UniProt/SwissProt Accession				
UNIF TOFOWISSE TOT ACCESSION				
Dataset				
[None Selected]				
[None Selected]				
😸 Done				Tor Disabled

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).