

## Module 5 – BioMart

### *You will learn about*

- BioMart, a joint project developed and maintained at EBI and OICR [www.biomart.org](http://www.biomart.org)
- How to use BioMart to quickly obtain lists of gene information from Ensembl
- Specifically, how to export a table of gene information in Microsoft Excel format, HTML or text, and/or sequence in Fasta format.

BioMart can be used to directly access the data in Ensembl and export tables of gene information or sequences. Any user can obtain gene-associated data in tabular format without the need for knowing any programming. The 'query' or the initial input can be an entire set of genes for a species, or a smaller more limited set (e.g. a list of IDs or a specific region of a chromosome). Information about the gene set defined by the user can be exported as txt, html, or in Microsoft Excel format (XLS). This information can range from chromosomal position to associated IDs in other databases to a short description of the gene. Other supported export formats are Fasta and **GFF**. These are only some examples of the information that can be obtained through this fast and user-friendly interface.

The following is a 'worked example' or web-site walkthrough of BioMart. It is probably the best way to learn how to use it! Read along, or follow on the web using the archive site for version 52 so that the layout is identical and results match up.

BioMart may have been updated since the time of this worked-example, but the concepts and basic layout should be the same. You can also find BioMart at [www.biomart.org](http://www.biomart.org) (click on MartView). Not only the Ensembl genes are accessible from BioMart; this tool can also be used to access data in **MSD** (Macromolecular Structures Database), **Wormbase**, **HapMap** and others.

### **BioMart Walkthrough:**

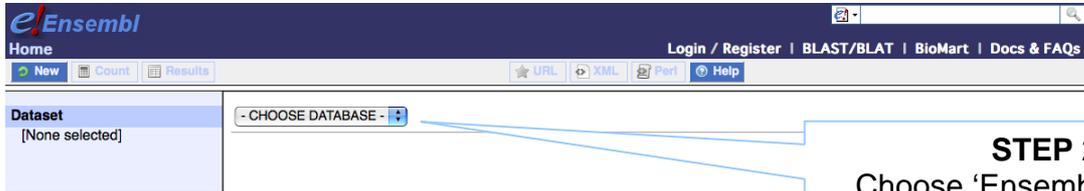
The human gene encoding Glucose-6-phosphate dehydrogenase (G6PD) is located on chromosome X in cytogenetic band q28.

Which other genes related to human diseases locate to the same band? What are their Ensembl Gene IDs and Entrez Gene IDs?

What are their cDNA sequences?

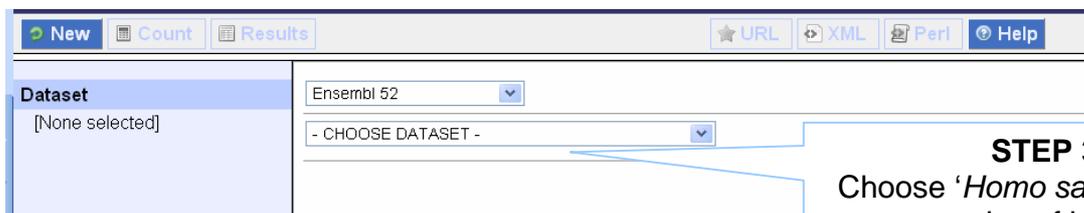
Follow the worked example below to answer these questions.

**Step 1:** Either click on 'BioMart' in the top right header bar of the Ensembl home page, or go to <http://www.biomart.org/> and click on the 'MartView' tab.



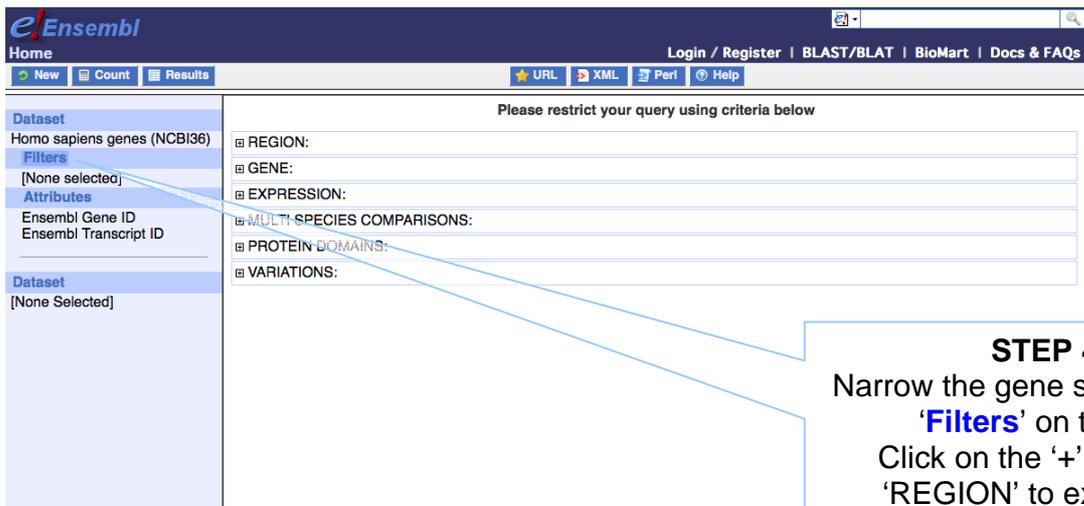
The screenshot shows the Ensembl 'New' query page. The 'Dataset' dropdown menu is currently set to '- CHOOSE DATABASE -'. A blue callout box points to this dropdown menu.

**STEP 2:**  
Choose 'Ensembl 52' as the primary database.



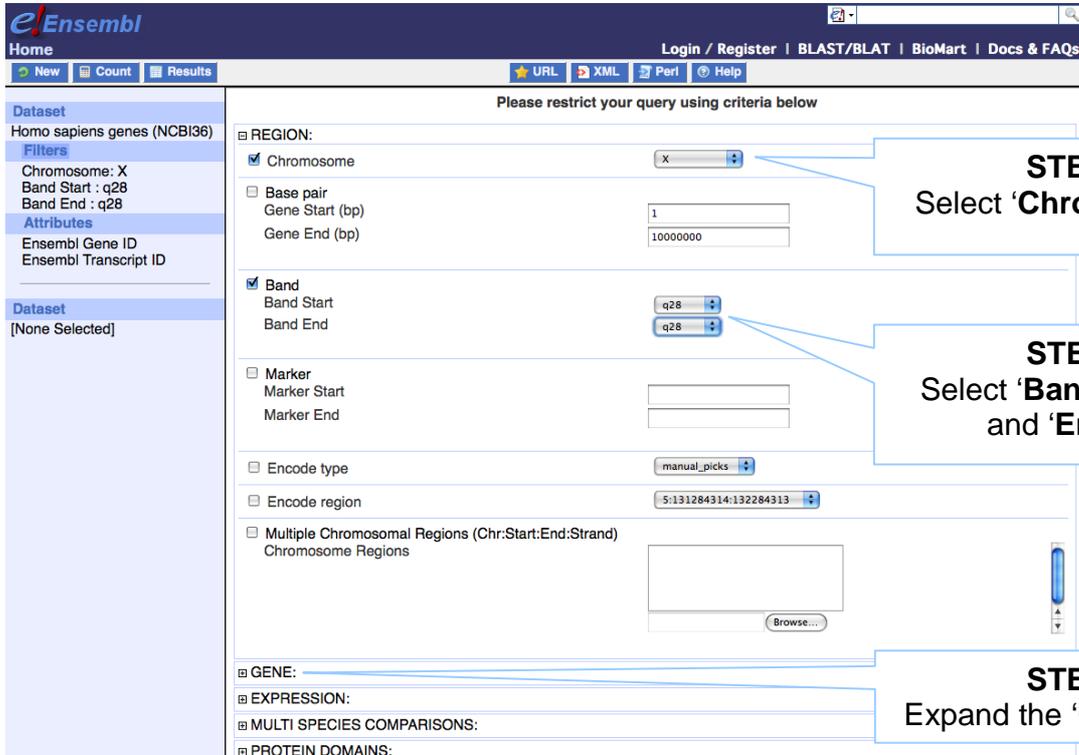
The screenshot shows the Ensembl 'New' query page. The 'Dataset' dropdown menu is now set to 'Ensembl 52'. A second dropdown menu below it is set to '- CHOOSE DATASET -'. A blue callout box points to this second dropdown menu.

**STEP 3:**  
Choose 'Homo sapiens' as the species of interest.



The screenshot shows the Ensembl 'New' query page with the 'Dataset' set to 'Homo sapiens genes (NCBI36)'. The 'Filters' section is expanded, showing options for 'REGION', 'GENE', 'EXPRESSION', 'MULTI-SPECIES COMPARISONS', 'PROTEIN DOMAINS', and 'VARIATIONS'. A blue callout box points to the 'Filters' section.

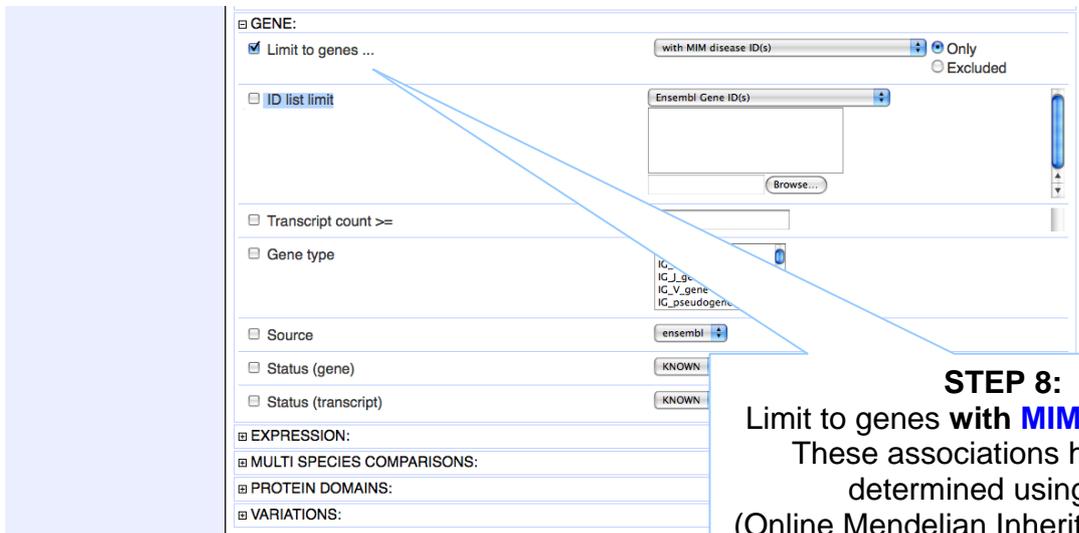
**STEP 4:**  
Narrow the gene set by clicking 'Filters' on the left. Click on the '+' in front of 'REGION' to expand the choices.



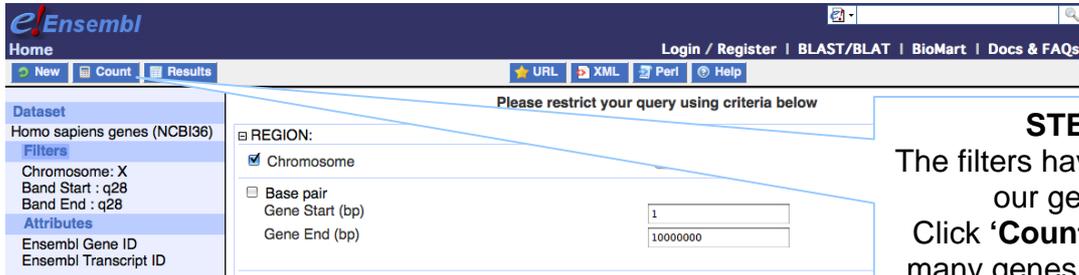
**STEP 5:**  
Select 'Chromosome X'

**STEP 6:**  
Select 'Band Start q28' and 'End q28'

**STEP 7:**  
Expand the 'GENE' panel.



**STEP 8:**  
Limit to genes with **MIM disease ID**.  
These associations have been determined using MIM (Online Mendelian Inheritance in Man).  
<http://www.ncbi.nlm.nih.gov/omim/>



Please restrict your query using criteria below

**Dataset**  
Homo sapiens genes (NCBI36)

**Filters**  
Chromosome: X  
Band Start : q28  
Band End : q28

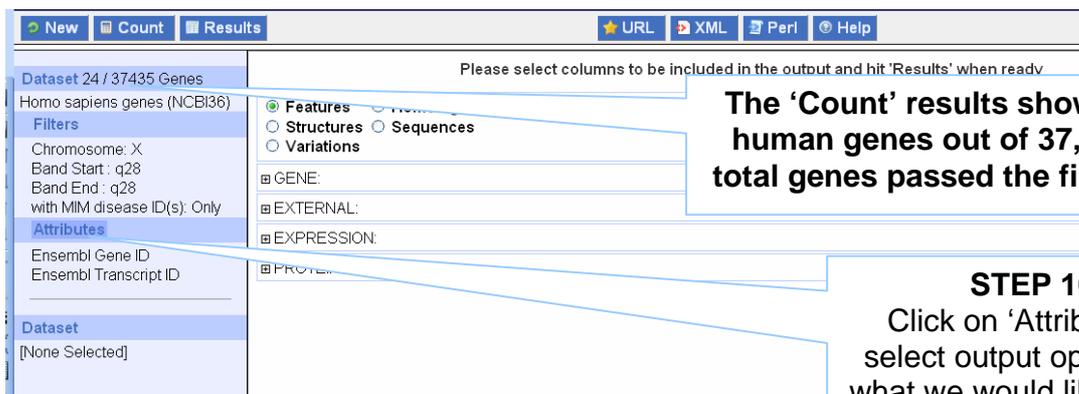
**Attributes**  
Ensembl Gene ID  
Ensembl Transcript ID

**REGION:**  
 Chromosome  
 Base pair  
Gene Start (bp)   
Gene End (bp)

**Count** | Results

URL | XML | Perl | Help

**STEP 9:**  
The filters have determined our gene set. Click 'Count' to see how many genes have passed these filters.



Please select columns to be included in the output and hit 'Results' when ready

**Dataset 24 / 37435 Genes**  
Homo sapiens genes (NCBI36)

**Filters**  
Chromosome: X  
Band Start : q28  
Band End : q28  
with MIM disease ID(s): Only

**Attributes**  
Ensembl Gene ID  
Ensembl Transcript ID

**Features**  
 Features  
 Structures  
 Sequences  
 Variations

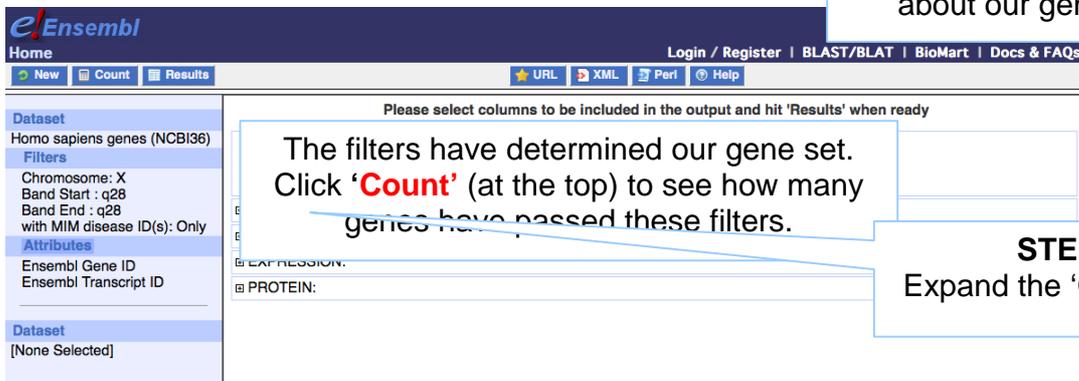
**GENE:**  
**EXTERNAL:**  
**EXPRESSION:**  
**PROTEIN:**

**Count** | Results

URL | XML | Perl | Help

The 'Count' results show 24 human genes out of 37,435 total genes passed the filters.

**STEP 10:**  
Click on 'Attributes' to select output options (i.e. what we would like to know about our gene set).



Please select columns to be included in the output and hit 'Results' when ready

**Dataset**  
Homo sapiens genes (NCBI36)

**Filters**  
Chromosome: X  
Band Start : q28  
Band End : q28  
with MIM disease ID(s): Only

**Attributes**  
Ensembl Gene ID  
Ensembl Transcript ID

**Features**  
 Features  
 Structures  
 Sequences  
 Variations

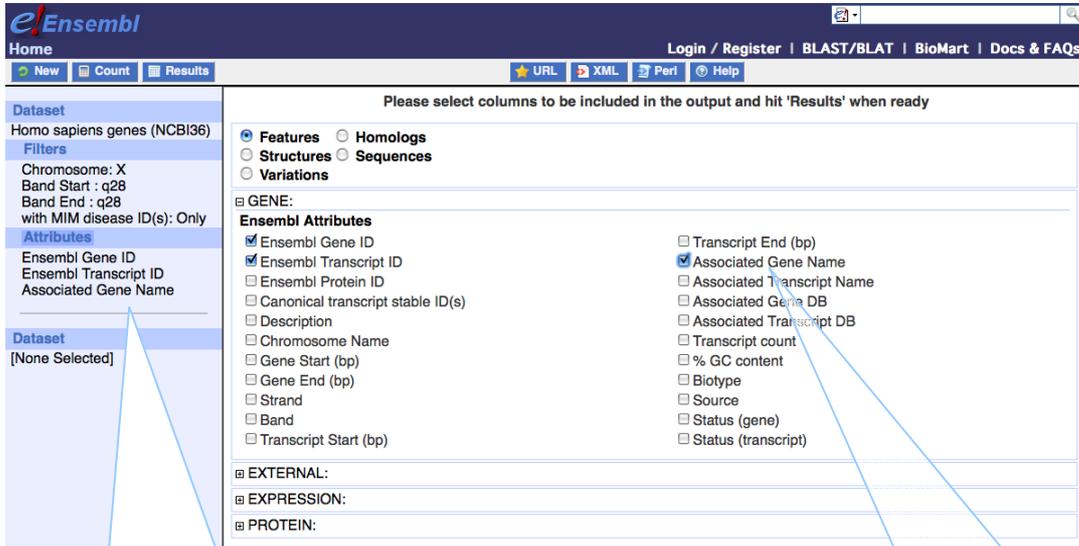
**GENE:**  
**EXTERNAL:**  
**EXPRESSION:**  
**PROTEIN:**

**Count** | Results

URL | XML | Perl | Help

The filters have determined our gene set. Click 'Count' (at the top) to see how many genes have passed these filters.

**STEP 11:**  
Expand the 'GENE' panel.



Home Login / Register | BLAST/BLAT | BioMart | Docs & FAQs

New Count Results URL XML Perl Help

**Dataset**  
Homo sapiens genes (NCBI36)

**Filters**  
Chromosome: X  
Band Start : q28  
Band End : q28  
with MIM disease ID(s): Only

**Attributes**  
Ensembl Gene ID  
Ensembl Transcript ID  
Associated Gene Name

**Dataset**  
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

Features  Homologs  
 Structures  Sequences  
 Variations

GENE:

**Ensembl Attributes**

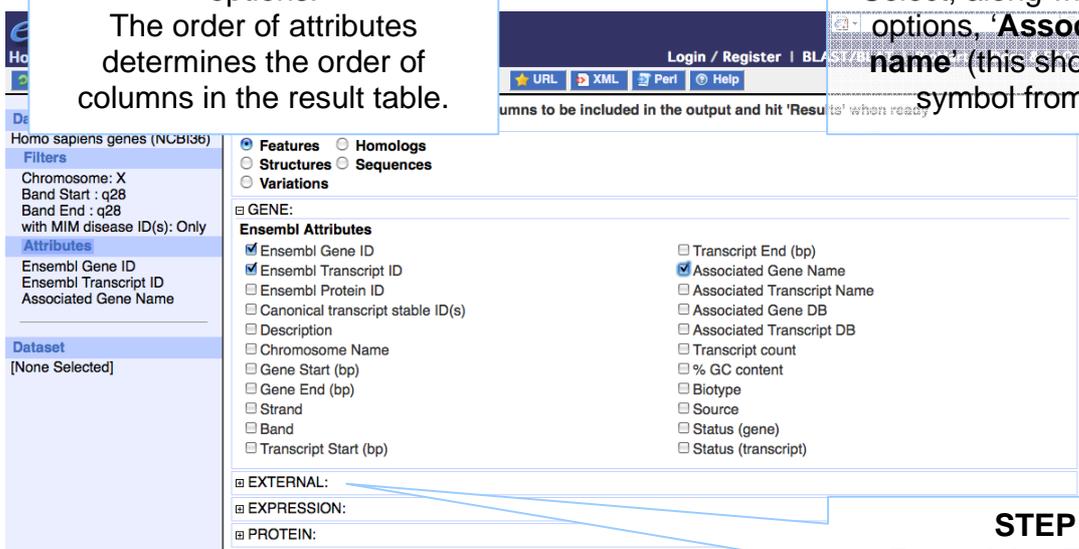
- Ensembl Gene ID
- Ensembl Transcript ID
- Ensembl Protein ID
- Canonical transcript stable ID(s)
- Description
- Chromosome Name
- Gene Start (bp)
- Gene End (bp)
- Strand
- Band
- Transcript Start (bp)

- Transcript End (bp)
- Associated Gene Name
- Associated Transcript Name
- Associated Gene DB
- Associated Transcript DB
- Transcript count
- % GC content
- Biotype
- Source
- Status (gene)
- Status (transcript)

EXTERNAL:  
 EXPRESSION:  
 PROTEIN:

Note the summary of selected options.  
The order of attributes determines the order of columns in the result table.

**STEP 12:**  
Select, along with the default options, 'Associated Gene name' (this shows the gene symbol from HGNC).



Home Login / Register | BLAST/BLAT | BioMart | Docs & FAQs

New Count Results URL XML Perl Help

**Dataset**  
Homo sapiens genes (NCBI36)

**Filters**  
Chromosome: X  
Band Start : q28  
Band End : q28  
with MIM disease ID(s): Only

**Attributes**  
Ensembl Gene ID  
Ensembl Transcript ID  
Associated Gene Name

**Dataset**  
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

Features  Homologs  
 Structures  Sequences  
 Variations

GENE:

**Ensembl Attributes**

- Ensembl Gene ID
- Ensembl Transcript ID
- Ensembl Protein ID
- Canonical transcript stable ID(s)
- Description
- Chromosome Name
- Gene Start (bp)
- Gene End (bp)
- Strand
- Band
- Transcript Start (bp)

- Transcript End (bp)
- Associated Gene Name
- Associated Transcript Name
- Associated Gene DB
- Associated Transcript DB
- Transcript count
- % GC content
- Biotype
- Source
- Status (gene)
- Status (transcript)

EXTERNAL:  
 EXPRESSION:  
 PROTEIN:

**STEP 13:**  
Expand the 'EXTERNAL' panel to select External References.

New Count Results URL XML Perl Help

Dataset 24 / 37435 Genes  
Homo sapiens genes (NCBI36)

Filters  
Chromosome: X  
Band Start: q28  
Band End: q28  
with MIM disease ID(s): Only

Attributes  
Ensembl Gene ID  
Ensembl Transcript ID  
Associated Gene Name  
EntrezGene ID  
MIM Morbid Accession  
MIM Morbid Description

External References (max 3)

- Clone based Ensembl gene name
- Clone based Ensembl transcript name
- Clone based VEGA gene name
- Clone based VEGA transcript name
- CCDS ID
- EMBL (Genbank) ID
- EntrezGene ID
- VEGA transcript ID(s)
- Ensembl transcript (where OTT is identical to OTT)
- HAVANA transcript (where ENST shares the same name as ENST)
- HAVANA transcript (where ENST identical to OTT)
- HGNC ID
- HGNC symbol
- HGNC automatic gene name
- HGNC curated gene name
- HGNC automatic transcript name
- HGNC curated transcript name
- PFI ID
- IMGT Gene DB
- MIM Morbid Accession
- MIM Morbid Description
- MIM Gene Accession
- MIM Gene Description
- miRBase
- PDB ID
- Protein ID
- RefSeq DNA ID
- RefSeq Predicted DNA ID
- RefSeq Protein ID
- RefSeq Predicted Protein ID
- Rfam ID
- UniProt
- UniProt/SwissProt
- UniProt Varsplice ID
- Human Protein Atlas
- Database of Aberrant Splicing
- DBASS3 Gene Name
- Database of Aberrant Splicing
- DBASS5 Gene Name

**STEP 14:**  
Select 'EntrezGene ID' and 'Mim Morbid Accession' and 'MIM Morbid Description'. These are MIM phenotypes and diseases, respectively.

New Count Results URL XML Perl Help

Home Login / Register | BLAST/BLAT | BioMart | Docs & FAQs

Please select columns to be included in the output and hit 'Results' when ready

Features Homologs  
Structures Sequences  
Variations

GENE:

Ensembl Attributes

- Ensembl Gene ID
- Ensembl Transcript ID
- Ensembl Protein ID
- Canonical transcript stable ID(s)
- Description
- Chromosome Name
- Gene Start (bp)
- Gene End (bp)
- Strand
- Band
- Transcript Start (bp)
- Transcript End (bp)
- Associated Gene Name
- Associated Transcript Name
- Associated Gene DB
- Associated Transcript DB
- Transcript count
- % GC content
- Biotype
- Source
- Status (gene)
- Status (transcript)

EXTERNAL:

**STEP 15:**  
Click 'RESULTS' at the top to preview the output.

New Count Results URL XML Perl Help

Dataset 24 / 37435 Genes  
Homo sapiens genes (NCBI36)

Filters  
Chromosome: X  
Band Start: q28  
Band End: q28  
with MIM disease ID(s): Only

Attributes  
Ensembl Gene ID  
Ensembl Transcript ID  
Associated Gene Name  
EntrezGene ID  
MIM Morbid Accession  
MIM Morbid Description

Export: all results to File TSV Unique results only

Email notification to

View 10 rows as HTML Unique results only

Ensembl Gene ID	Ensembl Transcript ID	Associated Gene Name	EntrezGene ID	MIM Morbid Accession	MIM Morbid Description
ENSG00000195010	ENST00000330256	FB	2157	316700	HEMOPHILIA A
ENSG00000195010	ENST00000330256	FB	2157	134500	FACTOR VIII DEFICIENCY
ENSG00000130626	ENST00000339550	DKFZ	1736	300240	HOYERHAAL-HREIDARSSON SYNDROME
ENSG00000130626	ENST00000339550	DKFZ	1736	305000	DYSKERATOSIS CONGENITA, X-LINKED
ENSG00000073009	ENST00000339602	IKBKG	8517	308300	INCONTINENTIA PIGMENTI
ENSG00000073009	ENST00000339602	IKBKG	8517	300640	INVASIVE PNEUMOCOCCAL DISEASE, RECURRENT ISOLATED, 2
ENSG00000073009	ENST00000339609	IKBKG	0517	300630	ATYPICAL MYCOBACTERIAL INFECTIONS, FAMILIAL, X-LINKED 1
ENSG00000073009	ENST00000339609	IKBKG	0517	300694	IMMUNODEFICIENCY WITH AUTOBROTIC ECTODERMAL DYSPLASIA
ENSG00000073009	ENST00000339609	IKBKG	8517	300301	ECTODERMAL DYSPLASIA ANHYDROTIC WITH IMMUNODEFICIENCY, OSTEOPETROSIS
ENSG00000073009	ENST00000339602	IKBKG	8517	300281	ECTODERMAL DYSPLASIA ANHYDROTIC WITH IMMUNE DEFICIENCY

**STEP 16:**  
Go back and change Filters or Attributes if desired.  
Or, View ALL rows as HTML...

To save a file of the complete table, click 'Go'. Or, email the results to any address.

Ensembl Gene ID	Ensembl Transcript ID	Associated Gene Name	EntrezGene ID	MIM Morbid Accession	MIM Morbid Description
<a href="#">ENSG00000185010</a>	<a href="#">ENST00000360256</a>	<a href="#">F8</a>	<a href="#">2157</a>	<a href="#">306700</a>	HEMOPHILIA A
<a href="#">ENSG00000185010</a>	<a href="#">ENST00000360256</a>	<a href="#">F8</a>	<a href="#">2157</a>	<a href="#">134500</a>	FACTOR VIII DEFICIENCY
<a href="#">ENSG00000130826</a>	<a href="#">ENST00000369550</a>	<a href="#">DKC1</a>	<a href="#">1736</a>	<a href="#">300240</a>	HOYERAAL-HREIDARSSON SYNDROME
<a href="#">ENSG00000130826</a>	<a href="#">ENST00000369550</a>	<a href="#">DKC1</a>	<a href="#">1736</a>	<a href="#">305000</a>	DYSKERATOSIS CONGENITA, X-LINKED
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369609</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">308300</a>	INCONTINENTIA PIGMENTI
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369609</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300640</a>	INVASIVE PNEUMOCOCCAL DISEASE, RECURRENT ISOLATED, 2
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369609</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300636</a>	ATYPICAL MYCOBACTERIOSIS, FAMILIAL, X-LINKED 1
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369609</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300584</a>	IMMUNODEFICIENCY WITHOUT ANHIDROTIC ECTODERMAL DYSPLASIA
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369609</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300301</a>	ECTODERMAL DYSPLASIA, ANHIDROTIC, WITH IMMUNODEFICIENCY, OSTEOPE TROSIS,
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369609</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300291</a>	ECTODERMAL DYSPLASIA, HYPOHIDROTIC, WITH IMMUNE DEFICIENCY
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369601</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">308300</a>	INCONTINENTIA PIGMENTI
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369601</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300640</a>	INVASIVE PNEUMOCOCCAL DISEASE, RECURRENT ISOLATED, 2
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369601</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300636</a>	ATYPICAL MYCOBACTERIOSIS, FAMILIAL, X-LINKED 1
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369601</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300584</a>	IMMUNODEFICIENCY WITHOUT ANHIDROTIC ECTODERMAL DYSPLASIA
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369601</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300301</a>	ECTODERMAL DYSPLASIA, ANHIDROTIC, WITH IMMUNODEFICIENCY, OSTEOPE TROSIS,
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369601</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300291</a>	ECTODERMAL DYSPLASIA, HYPOHIDROTIC, WITH IMMUNE DEFICIENCY
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369606</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">308300</a>	INCONTINENTIA PIGMENTI
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369606</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300640</a>	INVASIVE PNEUMOCOCCAL DISEASE, RECURRENT ISOLATED, 2
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369606</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300636</a>	ATYPICAL MYCOBACTERIOSIS, FAMILIAL, X-LINKED 1
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369606</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300584</a>	IMMUNODEFICIENCY WITHOUT ANHIDROTIC ECTODERMAL DYSPLASIA
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369606</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300301</a>	ECTODERMAL DYSPLASIA, ANHIDROTIC, WITH IMMUNODEFICIENCY, OSTEOPE TROSIS,
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369606</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300291</a>	ECTODERMAL DYSPLASIA, HYPOHIDROTIC, WITH IMMUNE DEFICIENCY
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369607</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">308300</a>	INCONTINENTIA PIGMENTI
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369607</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300640</a>	INVASIVE PNEUMOCOCCAL DISEASE, RECURRENT ISOLATED, 2
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369607</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300636</a>	ATYPICAL MYCOBACTERIOSIS, FAMILIAL, X-LINKED 1
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369607</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300584</a>	IMMUNODEFICIENCY WITHOUT ANHIDROTIC ECTODERMAL DYSPLASIA
<a href="#">ENSG00000073009</a>	<a href="#">ENST00000369607</a>	<a href="#">IKBK G</a>	<a href="#">8517</a>	<a href="#">300301</a>	ECTODERMAL DYSPLASIA, ANHIDROTIC, WITH IMMUNODEFICIENCY, OSTEOPE TROSIS,

Result Table 1

e!Ensembl

Home [Login / Register](#) | [BLAST/BLAT](#) | [BioMart](#) | [Docs & FAQs](#)

[New](#) | [Count](#) | [Results](#) | [URL](#) | [XML](#) | [Perl](#) | [Help](#)

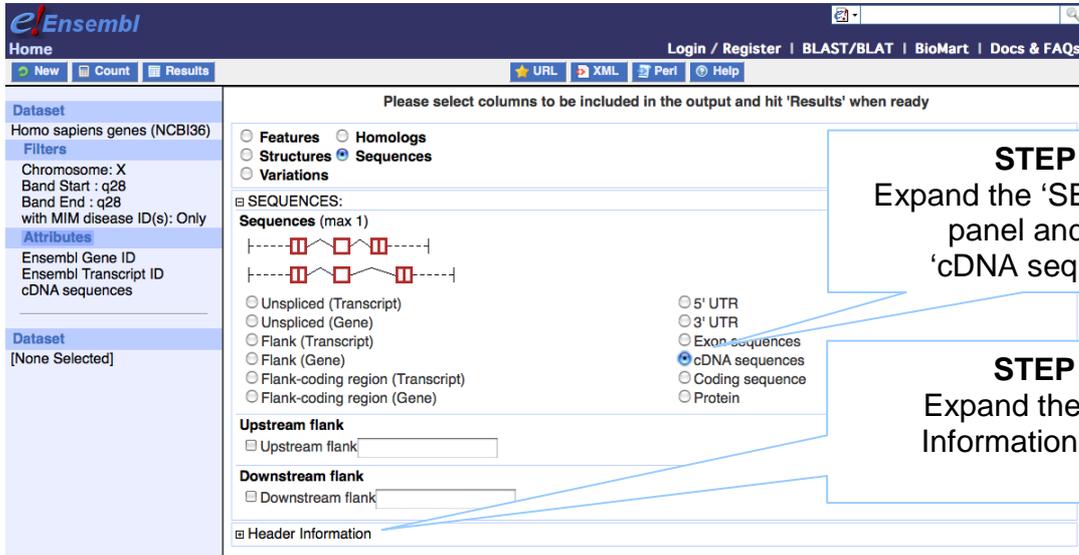
Please select columns to be included in the output and hit 'Results' when ready

Features
  Homologs
  Structures
  Sequences
  Variations

SEQUENCES:
  Header Information

**STEP 17:**  
 To view sequences, go back to 'Attributes'

**STEP 18:**  
 Select 'Sequences' and then expand the 'SEQUENCES' section.



Please select columns to be included in the output and hit 'Results' when ready

**STEP 19:**  
Expand the 'SEQUENCES' panel and select 'cDNA sequences'.

**STEP 20:**  
Expand the 'Header Information' section.

Dataset: Homo sapiens genes (NCBI36)  
Filters: Chromosome: X, Band Start: q28, Band End: q28, with MIM disease ID(s): Only  
Attributes: Ensembl Gene ID, Ensembl Transcript ID, cDNA sequences  
Dataset: [None Selected]

Features  Homologs  
 Structures  Sequences  
 Variations

SEQUENCES:

Sequences (max 1)

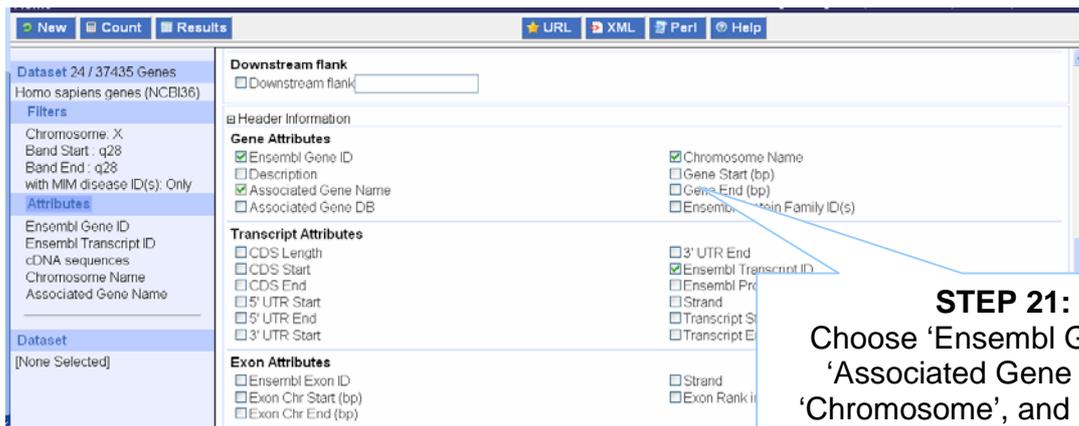
Unspliced (Transcript)  
 Unspliced (Gene)  
 Flank (Transcript)  
 Flank (Gene)  
 Flank-coding region (Transcript)  
 Flank-coding region (Gene)

5' UTR  
 3' UTR  
 Exon sequences  
 cDNA sequences  
 Coding sequence  
 Protein

Upstream flank  
 Upstream flank [ ]

Downstream flank  
 Downstream flank [ ]

Header Information



Dataset 24 / 37435 Genes  
Homo sapiens genes (NCBI36)  
Filters: Chromosome: X, Band Start: q28, Band End: q28, with MIM disease ID(s): Only  
Attributes: Ensembl Gene ID, Ensembl Transcript ID, cDNA sequences, Chromosome Name, Associated Gene Name  
Dataset: [None Selected]

Downstream flank  
 Downstream flank [ ]

Header Information

Gene Attributes

Ensembl Gene ID  
 Description  
 Associated Gene Name  
 Associated Gene DB  
 Chromosome Name  
 Gene Start (bp)  
 Gene End (bp)  
 Ensembl Protein Family ID(s)

Transcript Attributes

CDS Length  
 CDS Start  
 CDS End  
 5' UTR Start  
 5' UTR End  
 3' UTR Start  
 3' UTR End  
 Ensembl Transcript ID  
 Ensembl Protein ID  
 Strand  
 Transcript Start  
 Transcript End

Exon Attributes

Ensembl Exon ID  
 Exon Chr Start (bp)  
 Exon Chr End (bp)  
 Strand  
 Exon Rank

**STEP 21:**  
Choose 'Ensembl Gene ID', 'Associated Gene Name', 'Chromosome', and 'Ensembl Transcript ID'

Home Login / Register | BLAST/BLAT | BioMart | Docs & FAQs

New Count Results URL XML Perl Help

**Dataset**

Homo sapiens genes (NCBI36)

**Filters**

Chromosome: X  
Band Start : q28  
Band End : q28  
with MIM disease ID(s): Only

**Attributes**

Ensembl Gene ID  
Ensembl Transcript ID  
cDNA sequences

---

**Dataset**

[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

Features    Homologs  
 Structures    Sequences  
 Variations

**SEQUENCES:**

**Sequences (max 1)**

Unspliced (Transcript)  
 Unspliced (Gene)  
 Flank (Transcript)  
 Flank (Gene)  
 Flank-coding region (Transcript)  
 Flank-coding region (Gene)

5' UTR  
 3' UTR  
 Exon sequences  
 cDNA sequences  
 Coding sequence  
 Protein

**Upstream flank**

Upstream flank

**Downstream flank**

Downstream flank

Header Information

**STEP 22:**  
Click 'Results'

New Count Results URL XML Perl Help

**Dataset 24 / 37435 Genes**

Homo sapiens genes (NCBI36)

**Filters**

Chromosome: X  
Band Start : q28  
Band End : q28  
with MIM disease ID(s): Only

**Attributes**

Ensembl Gene ID  
Ensembl Transcript ID  
cDNA sequences  
Chromosome Name  
Associated Gene Name

---

**Dataset**

[None Selected]

Export all results to  FASTA  Unique results only

Email notification to

View  rows as   Unique results only

```

> ENSG00000130821|ENST00000253122|X|SLC6A8
TAGTGGAGCGAGGTGCGAGTGCCTGAGCCGCGCGCGCCCGAGAGCGGCTGCAGCCG
CCGCGCCCGGGAAGGAGAGGCGAGGCGCGCGCGCGCCCGCCCGCCCGCCCGCCCGCC
CCGCGCGCACCAACCGCCACCGAGTCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
CCGATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
CTCACTTTCAAGTGTCTTCTGAGCTGCTTGAATGCGCGCGCGCGCGCGCGCGCGCGCG
CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
CCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
ACGGCATCTATAGCGTGTCCGCGCGAGGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCG
ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
CGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
GTGTGTCCTTATCCCTACGTCCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
.....
                    
```

Again, View ALL rows as FASTA for the full list... (make sure pop-up moduleer is off).



## RESULTS

>Header: Gene ID, Transcript ID, Chromosome and Gene Name

```
>ENSG00000073009 | ENST00000369601 | X | IKBK
AGCCCGTTCCTGCTCCGCGCTTCTGGAGCACTGGCCAAGCGGGCCGATTCCAGGACCCAG
GTTACTTGGGGCGGCGAGCTGGACTGTTTCTACTCCTCCCTCCCTCCACTGCGGGGTCT
GACCCCTACTCCTTGTGTGAGGACTCCTCTAGTTTCAGAGACATATTCTGTTCACAAAAC
GACTGGCTCTATCGAGGTGTTAAATTTCTCGGAAATGCCTCACATATAGTTTGGCAGC
TAGCCCTTGCCTTGTGGATGAATAGGCACCTCTGGAAGAGCCAACTGTGTGAGATGGTG
CAGCCCAGTGGTGGCCCGGAGCAGATCAGGACGTACTGGGCGAAGAGTCTCCTCTGGGG
AAGCCAGCCATGCTGCACCTGCCTTCAGAACAGGGCGCTCCTGAGACCCTCCAGCGCTGC
CTGGAGGAGAAATCAAGAGCTCCGAGATGCCATCCGGCAGAGCAACCAGATTCTGCGGGAG
CGCTGCGAGGAGCTTCTGCATTTTCCAAGCCAGCCAGAGGGAGGAGAAAGGAGTTTCTCATG
TGCAAGTTCCAGGAGGCCAGGAAACTGGTGGAGAGACTCGGCCCTGGAGAAGCTCGATCTG
AAGAGGCAGAGGAGCAGGCTCTGCGGGAGGTGGAGCACCTGAAGAGATGCCAGCAGCAG
ATGGCTGAGGACAAGGCCCTCTGTGAAAGCCCAGGTGACGTCTTGTCTCGGGGAGCTGCAG
GAGAGCCAGAGTCCGTTGGAGGCTGCCACTAAGGAATGCCAGGCTCTGGAGGGTTCGGGCC
CGGGCGGCCAGCAGCAGGCGCGGAGCTGGAGAGTGAGCGGAGGCGCTGCAGCAGCAG
CACAGCGTGCAGGTGGACCAGCTGCGCATGCAGGGCCAGAGCGTGGAGGCCGCGCTCCGC
ATGGAGCGCCAGGGCGCCTCGGAGGAGAAAGGAAAGCTGGCCAGTTGCAGGTGGCCTAT
CACCAGCTCTTCCAAGAATACGACAACCACATCAAGAGCAGCGTGGTGGCCAGTGCAGCGG
AAGCCAGGAAATGCAGCTGGAAGATCTCAAACAGCAGCTCCAGCAGGCCGAGGAGGCCCTG
GTGGCCAAACAGGAGGTGATCGATAAGCTGAAGGAGGAGGCCGAGCAGCACAAAGATTGTG
ATGGAGACCGTTCCGGTGTGAAAGGCCAGGCGGATATCTACAAAGCCGACTTCCAGGCT
GAGAGGCAGGCCCGGGAGAAAGCTGGCCGAGAAGAAGGAGCTCCTGCAGGAGCAGCTGGAG
CAGCTGCAGAGGGAGTACAGCAAACTGAAGGCCAGCTGTCAGGAGTCCGCCAGGATCCGAG
GACATGAGGAAGCGGCATGTCAGGTTCTCCAGGCCCTTGGCCCCCGCCCTGCCTAC
CTCTCTCTCCCTGGCCCTGCCAGCCAGAGGAGGAGCCCCCGAGGAGCCACCTGAC
TTCTGCTGTCCCAAGTGCCAGTATCAGGCCCTGATATGGACACCCTGCAGATACATGTC
ATGGAGTGCATTGAGTAGGGCCGGCCAGTGCAAGGCCACTGCCTGCCGAGGACGTGCCCG
GGACCGTGCAGTCTGCGCTTTCTCTCCCGCCTGCCTAGCCCAAGATGAAGGGCTGGGTG
GCCACAACCTGGGATGCCACCTGGAGCCCCACCCAGGAGCTGGCCGCGGCACCTTACGCTT
CAGCTGTTGATCCGCTGGTCCCTCTTTTGGGGTAGATGCGGCCCGGATCAGGCCTGACT
CGCTGCTCTTTTGTTCCTTCTGTCTGCTCGAACCACTTGCCCTCGGGCTAAATCCCTCCC
TCTTCTCCACCCCGCACTGGGGAAGTCAAGAATGGGGCTGGGGCTCTCAGGGAGAACT
GCTTCCCTGGCAGAGCTGGGTGGCAGCTTCTCTCCACCCGACACCCGACCCCGCCGCT
GCTGTGCCCCGGGAGTGTGCCCTCTTACCATGCACACGGGTGCTCTCTTTTGGGCTGC
ATGCTATTCCATTTTGCAGCCAGACCGATGTGTATTTAACAGTCACTATTGATGGACAT
TTGGGTTGTTTCCATCTTTTGTACCATAAATAATGGCATAGTAAAAATCCTTGTGCA
TT
```

cDNA 1

```
>ENSG00000126895 | ENST00000358927 | X | AVPR2
TTCACGCCACCGCCAGCTGCCAGGAGCCAGCCAGGACTGGCCATACTGCCACCCGACA
CGTGACACACCGCCAACAGGCATCTGCCATGCTGGCATCTCTATAAGGGCTCCAGTCCAG
AGACCTGGGCCATTGAACCTTGTCTCCTCAGGCAGAGGCTGAGTCCGCACATCACCTCCAG
GCCCTCAGAACACCTGCCCCAGCCCCACCATGCTCATGGCGTCCACCCTTCCGCTGTGC
CTGGGCATCCCTCTCTGCCAGCCTGCCAGCAACAGCAGCCAGGAGAGGCCACTGGACA
CCCGGGACCCGCTGCTAGCCCGGGCGGAGCTGGCGCTGCTCTCCATAGTCTTTGTGGCTG
TGGCCCTGAGCAATGGCCTGGTGTGGCGGCCCTAGCTCGGCCGGGCGGGCGGGCCACT
GGGCACCCATACACGTCTTTCATTGGCCACTTGTGCTGCGCCGACCTGGCCGCTGGCTCTGI
TCCAAGTGTGCCCCAGCTGGCCTGGAAGGCCACCGACCGCTTCCGTGGGCCAGATGCC
TGTGTCCGGCGGTGAAGTATCTGCAGATGGTGGGCATGTATGCTCCTCCTACATGATCC
TGGCCATGACCGCTGCAAGCCACCGCTGCAATCTGCGCTGCCATGCTGGCCTACCGCATG
```

cDNA 2

## Glossary

**Attributes** (*In BioMart*) Information to attach to the geneset selected, either sequence or column headers

**Biotype** (*In BioMart*) Gene type (i.e. coding, non-coding)

**Entrez Gene** NCBI searchable database of gene sequences. Accession numbers in GenBank agree with DDBJ and EMBL.

**External Gene ID** An ID or accession number in a database apart from Ensembl

**Filters** (*In BioMart*) Information applied to narrow the selection, such as filtering the entire geneset for a species down to only genes on a specific chromosomal region with a GO term or Interpro domain. IDs can be used as 'filters' to select a gene set by an ID list.

**GFF** A file format often used in genetics, applicable across programs and databases

**HapMap** An international partnership committed to the development of a haplotype map describing common patterns of the human genome. <http://www.hapmap.org/>

**HGNC – The HUGO Gene Nomenclature Committee** A committee focused on the determination of one unique symbol for every human gene. [www.genenames.org/](http://www.genenames.org/)

**InterPro** A database of common protein motifs and domains, accesses information across a large number of protein databases. <http://www.ebi.ac.uk/interpro/>

**OMIM** 'Online Mendelian Inheritance in Man'. A database of phenotypic information for human. <http://www.ncbi.nlm.nih.gov/omim/>

**MSD** 'Macromolecular Structure Database' A collection of protein and other macromolecular structures, in part from the PDB (Protein Data Bank). <http://www.ebi.ac.uk/msd/>

**WormBase** An in-depth look at the *C. elegans* genome and other worm genomes. A manually curated gene set, genome browser and WormMart are available. <http://www.wormbase.org/>

## What to do next

Watch the related video!

[www.ensembl.org/info/website/tutorials/index.html](http://www.ensembl.org/info/website/tutorials/index.html)

'Introduction to BioMart'