

## BioMart

### Mining data- worked example

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 image shows two screenshots of the Ensembl BioMart interface. The first screenshot shows the 'Dataset' section with a dropdown menu set to '- CHOOSE DATABASE -'. A callout box labeled 'STEP 2:' points to this dropdown with the instruction: 'Choose 'Ensembl 52' as the primary database.' The second screenshot shows the 'Dataset' section with the dropdown menu set to 'Ensembl 52'. A callout box labeled 'STEP 3:' points to a second dropdown menu labeled '- CHOOSE DATASET -' with the instruction: 'Choose 'Homo sapiens' as the species of interest.'

**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/>

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

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).

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

Ensembl Home

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

Features
  Homologs  
 Structures
  Sequences  
 Variations

GENE:  
 EXTERNAL:  
 EXPRESSION:  
 PROTEIN:

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

Ensembl Home

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

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

Ensembl Home

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.

**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) (VEGA)
- Ensembl transcript (where OTT is identical to OTTT)
- HAVANA transcript (where ENST shares the same name with ENST)
- HAVANA transcript (where ENST identical to OTTT)
- HGNC ID
- HGNC symbol
- HGNC automatic gene name
- HGNC curated gene name
- HGNC automatic transcript name
- HGNC curated transcript name
- IPI ID
- IMGT Gene DB
- EMBL/TrEMBL/UniProt
- 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 II
- 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.

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

**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)

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

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
ENSG00000189010	ENST00000260026	FB	2157	316700	HEMOPHILIA A
ENSG00000189010	ENST00000260026	FB	2157	134500	FACTOR VIII DEFICIENCY
ENSG00000130826	ENST00000289950	DIK1	1736	300240	HOYERAAAL-HREIDARSSON SYNDROME
ENSG00000130826	ENST00000289950	DIK1	1736	305000	DYSKERATOSIS CONGENITA, X-LINKED
ENSG00000073009	ENST00000289609	IKBK1	8517	308300	INCONTINENTIA PIGMENTI
ENSG00000073009	ENST00000289609	IKBK1	8517	300640	INVASIVE PNEUMOCOCCAL DISEASE, RECURRENT ISOLATED, 2
ENSG00000073009	ENST00000289609	IKBK1	8517	300630	ATYPICAL MYCOBACTERIOSIS, FAMILIAL, X-LINKED 1
ENSG00000073009	ENST00000289609	IKBK1	8517	300584	IMMUNODEFICIENCY WITH AIR ANHIDROTIC ECTODERMAL DYSPLASIA
ENSG00000073009	ENST00000289609	IKBK1	8517	300301	ECTODERMAL DYSPLASIA ANHIDROTIC WITH IMMUNODEFICIENCY, OSTEOPEETROSIS
ENSG00000073009	ENST00000289609	IKBK1	8517	300291	ECTODERMAL DYSPLASIA ANHIDROTIC, 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**

The screenshot shows the Ensembl BioMart interface. On the left, the 'Attributes' section is expanded, showing 'Protein', 'Ensembl Gene ID', and 'Ensembl Transcript ID'. In the main area, the 'Sequences' radio button is selected under the 'Please select columns to be included in the output and hit 'Results' when ready' heading. Two callout boxes are present: one pointing to the 'Attributes' section with the text 'STEP 17: To view sequences, go back to 'Attributes'', and another pointing to the 'Sequences' section with the text '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

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:   
Downstream flank:   
Header Information

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

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

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

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

**STEP 22:**  
Click 'Results'

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

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:   
Downstream flank:   
Header Information

[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

```

>ENS00000130821|ENST00000253122|X|SLC6A8
TAGTCGGAGCGAGGTGGCGAGTCGCTGAGCCCGCGGCCCCGAGAGCGGGCTGCAGCCG
CCGCCCCCGGAAAGGAGAGGGCGAGGCGCGCCCGAGCCCGCCCGCCCGCCCGCCCG
CCGCGCCACCAACCGCCACCGGAGTCGCGGGCCAGCCGGGACGCTCCGCGGGCCCCGGC
CGGGCGGGGGCGCGGGCCACAGGCCCTGCTCCGGCCCGCCCTTGCAGACCCCGGGCG
CGATGTCCCGCGCCCGCTAGGCTGAGCCTCGGGTCGGGCGAGGAGCCGCGCAGCC
GCCCGCCCGAGCCCGGGCAGGAGCTCGGGAGCCCGCCCGCCCGCCCGCCCGCCCG
GCCGGCCCCCGCCGCCCGCGCGCCCGGGCCCCGACACACATGAGATTCTTCAAG
CTCACTTCAAGTCTTCGTGGACTGCTTCTGACTGCGCCCGCCCGCCCGCCCGCCCG
CGCCCGCCCGCCCGGTCGCCCGGCCCGGCCCGCCCGGCCCGCCCGGCCCGCCCGCC
CCTCGGGCCCTCCCGGTGCCCGGTGCCCGCCCGCTGACCCCGCCCGCCCGGTGAGGC
GCCGACCCCGCCCGCGGTGCGGCCCGCCGAGGCCATGCGGAAAGAGGCCCGGAG
ACGGCATCTATAGCGTGTCCGGCGACGAGAAAGAGGCCCCCTCATCGCGCCGGGCCG
ACGGGCCCGCCGCAAGGGCGACGCGCCCGTGGCCCTGGGACACCCCGCGCCCGCTGG
CCGTGCCCGCGCGAGACTGGACGCGCCAGATGGACTTCAATGTCGTGCGTGGGCT
TCGCGTGGGCTTGGCAACGTGTGGCGCTTCCCTACCTGTGCTACAAGAACGGCGAG
GTGTGTCTTATTCCCTACGTCTGATCGCCCTGGTTGGAGGAATCCCGATTTCTTCT
  
```

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

## RESULTS

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

```
>ENSG00000073009|ENST00000369601|X|IKBK1
AGCCCGTTCTGCTCCGCGCTTCTGGAGCACTGGCCAAGGCGGGCCGATTTCAGGACCCAG
GTTACTTGGGGGGGAGCTGGACTGTTTCTACTCCTCCCTCCTCCTCCACTGCGGGGTCT
GACCCCTACTCCTTGTGTGAGGACTCCTCTAGTTTCAGAGACATATTCTGTTCCACAAACTT
GACTGCGCTCTATCGAGGTCGTTAAATTTCTTCGGAATGCCTCACATATAGTTTGGCAGC
TAGCCCTTGCCCTGTGGATGAATAGGCACCTCTGGAAGAGCCAACCTGTGTGAGATGGTG
CAGCCCAGTGGTGGCCCGGAGCAGATCAGGACGTACTGGGCGAAGAGTCTCCTCTGGGG
AAGCCAGCCATGCTGCACCTGCCTTCAGAACAGGGCGCTCCTGAGACCCTCCAGCGCTGC
CTGGAGGAGAATCAAGAGCTCCGAGATGCCATCCGGCAGAGCAACCAGATTCTGCGGGAG
CGCTGCGAGGAGCTTCTGCATTTCCAAGCCAGCCAGAGGGAGGAGAAGGAGTCTCTCATG
TGCAAGTTCCAGGAGGCCAGGAAACTGGTGGAGAGACTCGGCCCTGGAGAAGCTCGATCTG
AAGAGGCAGAAAGGAGCAGGCTCTGCGGGAGGTGGAGCACCTGAAAGAGATGCCAGCAGCAG
ATGGCTGAGGACAAGGCCCTCTGTGAAAGCCAGGTGACGTCTTGTCTGCGGGGAGCTGCAG
GAGAGCCAGAGTCGCTTGGAGGCTGCCACTAAGGAATGCCAGGCTCTGGAGGGTCCGGCC
CGGGCGGCCAGCGAGCGCCAGGCAAGCTGGAGAGTGAAGCGGAGGCGCTGCAGCAGCAG
CACAGCGTGCAGGTGGACCAGCTGCGCATGCAGGGCCAGAGCGTGGAGGCCCGCTCCGC
ATGGAGCGCCAGGCCCGCTCGGAGGAGAAGAGGAAGCTGGCCAGTTGCAGGTGGCCTAT
CACCAGCTCTTCCAAGAATACGACAACCACATCAAGAGCAGCGTGGTGGGCAGTGAAGCGG
AAGCGAGGAATGCAGCTGGAAAGATCTCAAACAGCAGCTCCAGCAGGCCGAGGAGGCCCTG
GTGGCCAAACAGGAGGTGATCGATAAAGCTGAAAGGAGGAGGCCGAGCAGCACAAAGATTGTG
ATGGAGACCCTTCCGGTGTGAAGGCCAGGCCGATATCTACAAGGCCGACTTCCAGGCT
GAGAGGCAGGCCCGGGAGAGCTGGCCGAGAAGAAGGAGCTCCTGCAGGAGCAGCTGGAG
CAGCTGCAGAGGGAGTACAGCAAACCTGAAGGCCAGCTGTCAGGAGTCCGGCCAGGATCGAG
GACATGAGGAAGCGGCATGTCTGAGGTCTCCAGGCCCTTGGCCCCCGCCCTGCCTAC
CTCTCTCTCCCTGGCCCTGCCAGCCAGAGGAGGAGCCCCCGAGGAGCCACCTGAC
TTCTGCTGTCCCAAGTGCCAGTATCAGGCCCTTGATATGGACACCCTGCAGATACATGTC
ATGGAGTGCATTGAGTAGGGCCGGCCAGTGCAAGGCCACTGCCTGCCGAGGACGTGCCCG
GGACCGTGCAGTCTGCGCTTTCCTCTCCCGCCTGCTAGCCAGGATGAAGGGCTGGGTG
GCCACAACCTGGGATGCCACCTGGAGCCCCACCCAGGAGCTGGCCGCGGCACCTTACGCTT
CAGCTGTTGATCCGCTGGTCCCTCTTTTGGGGTAGATGCGGCCCGATCAGGCCTGACT
CGCTGCTCTTTTTGTTCCCTTCTGTCTGCTCGAACCACTTGCCCTCGGGCTAATCCCTCCC
TCTTCTCCACCCGGCACTGGGGAAAGTCAAGAATGGGGCCTGGGGCTCTCAGGGAGAAT
GCTTCCCTGGCAGAGCTGGGTGGCAGCTCTTCTCCACCCGACACCCGACCCGCCCCGCT
GCTGTGCCCTGGGAGTGTGCCCCCTTACCATGCACACGGGTGCTCTCCTTTTGGGCTGC
ATGCTATTCCATTTTGCAGCCAGACCGATGTGTATTTAAACCGTCACTATTGATGGACAT
TTGGGTTGTTTCCATCTTTTGTACCATAAATAATGGCATAGTAAAAATCCTTGTGCA
TT
```

cDNA 1

```
>ENSG00000126895|ENST00000358927|X|AVPR2
TTCACGCCACCGCCAGCTGCCAGGAGCCAGCCAGGACTGGCCATACTGCCACCGACA
CGTGACACACCGCCAACAGGCATCTGCCATGCTGGCATCTCTATAAGGGCTCCAGTCCAG
AGACCTGGGCCATTGAACTTGTCTCTCAGGCAGAGGCTGAGTCCGCACATCACCTCCAG
GCCCTCAGAACACCTGCCCCAGCCCCACCATGCTCATGGCGTCCACCACTTCCGCTGTGC
CTGGGCATCCCTCTCTGCCAGCCTGCCAGCAACAGCAGCCAGGAGAGGCCACTGGACA
CCCCGGACCCGCTGCTAGCCCCGGGCGGAGCTGGCGCTGCTCTCCATAGTCTTTGTGGCTG
TGGCCCTGAGCAATGGCCTGGTGTGGCGGCCCTAGCTCGGGCGGGCCGGCGGGGCCACT
GGCACCCATACAGTCTTCAATTGGCCACTTGTGCCTGGCCGACCTGGCCGTGGCTCTGT
TCCAAGTGTGCCCCAGCTGGCCCTGGAAGGCCACCGACCGCTTCCGTGGGCCAGATGCC
TGTGTGGGGCGTGAAGTATCTGCAGATGGTGGGCATGTATGCCTCCTCCTACATGATCC
TGGCATGAGGCTGAAACCGCACCGTGGCATCTGCGCTGCCATGCTGGCGTACCGGCATG
```

cDNA 2