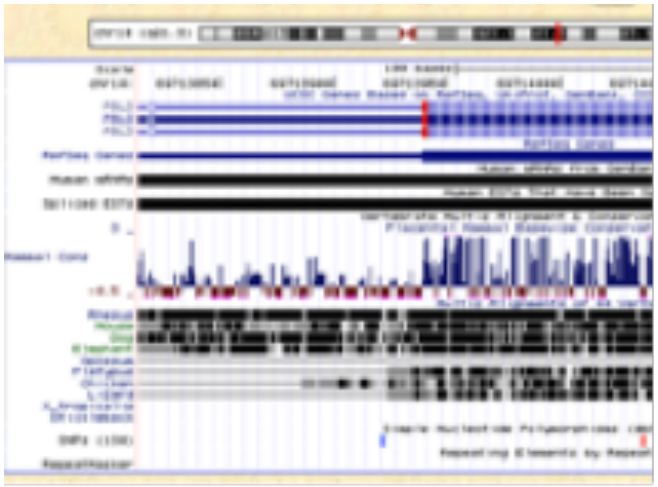
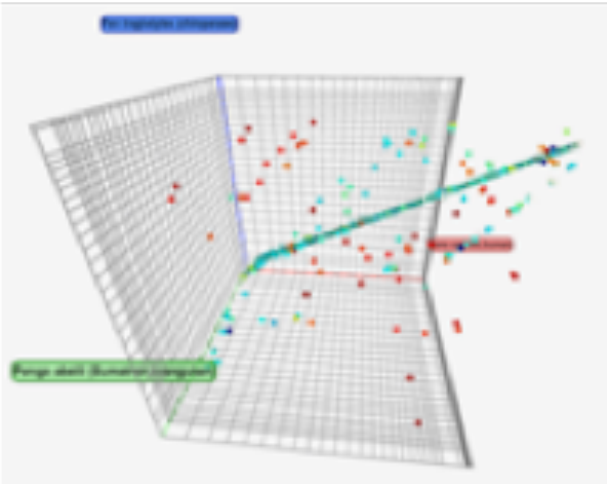
Computational Genomics

Data Retrieval II





Data Retrieval
Retrieving Data From ENSEMBL

FTP

Show 10 entries | Show/hide columns | Filter

★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Other annotations	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig
Y	**Human** *Homo sapiens*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
Y	**Mouse** *Mus musculus*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
Y	**Zebrafish** *Danio rerio*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	GVF	VCF	VEP	-	-	BAM/BigWig
	Abingdon island giant tortoise *Chelonoidis abingdonii*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	BAM/BigWig
	African ostrich *Struthio camelus australis*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	BAM/BigWig
	Agassiz's desert tortoise *Gopherus agassizii*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	BAM/BigWig
	Algerian mouse *Mus spretus*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	-
	Alpaca *Vicugna pacos*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	-
	Alpine marmot *Marmota marmota marmota*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	BAM/BigWig
	Amazon molly *Poecilia formosa*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	BAM/BigWig

Showing 1 to 10 of 311 entries | << < 1 2 3 4 5 > >>

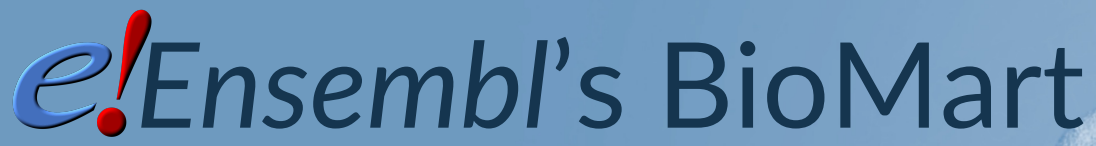
Data Retrieval
Retrieving Data From ENSEMBL

BioMart

EMBL-EBI

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining



EMBL-EBI

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

What is BioMart?

A tool in your browser to:

- Export data with no programming required
- Build queries with a few mouse clicks
- Generate custom data tables and sequence files

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

Training materials

- Ensembl training materials are protected by a CC BY license:
  creativecommons.org/licenses/by/4.0/
- If you wish to re-use these materials, please credit Ensembl for their creation
- If you use Ensembl for your work, please cite our papers:
  ensembl.org/info/about/publications.html

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

EMBL-EBI

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

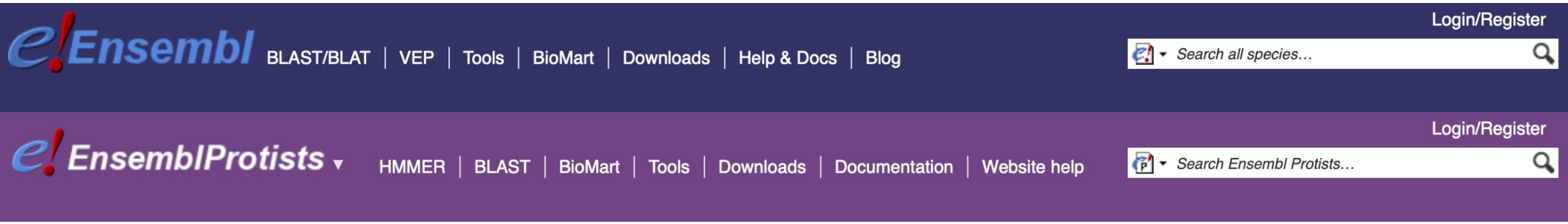
Why use BioMart?

For things that would be time consuming/difficult with the Ensembl browser:

- Query multiple things at once:
  - ID conversions
  - Gene locations
  - Download sequences
- Export large amounts of data

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining



Where can I find BioMart?

ensembl.org/biomart/martview

protists.ensembl.org/biomart/martview

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

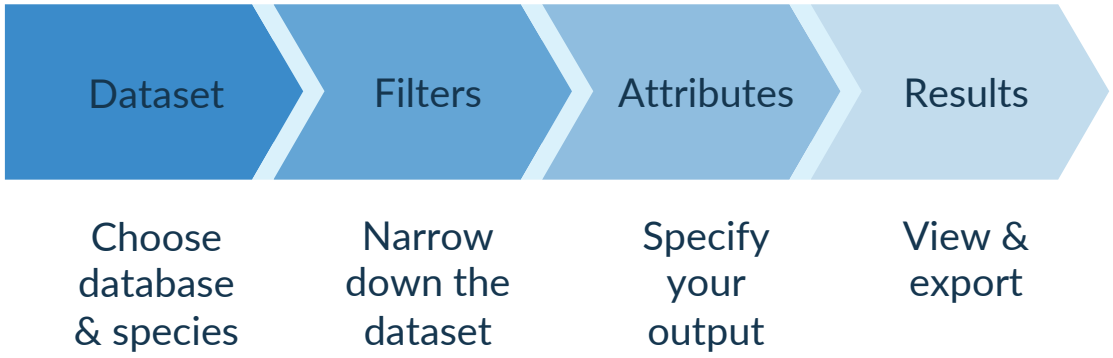
For which genomes is BioMart available?

- *Ensembl*
  (some exceptions)
- *EnsemblFungi*
  (some exceptions)
- *EnsemblMetazoa*
  (some exceptions)
- *EnsemblPlants*
  (some exceptions)
- *EnsemblProtists*
  (some exceptions)

Data Retrieval
Retrieving Data From ENSEMBL

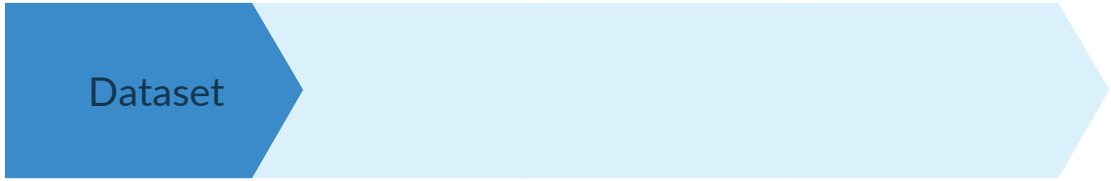
BioMart for Data Mining

How do I use BioMart?
The four steps



Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining



How do I use BioMart?
1. Dataset

- Define the database you want to search with your filters:
  - Genes
  - Variation
  - Regulation
  - Mouse strains (genes)
- Define the species

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

How do I use BioMart?
2. Filters

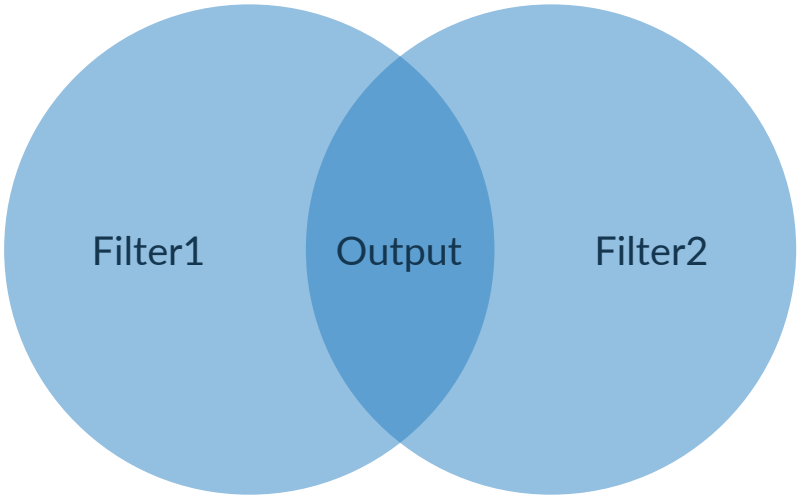
Define a (large) set of genes/variants by combination of filters, e.g.:

- Region
- List of IDs
- Function (GO term)
- Phenotype

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

How do I use BioMart?
2. Filters



Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

How do I use BioMart?
3. Attributes

Define the data you want to export (your output), e.g.:

- IDs
- Features
- Variants
- Orthologues/paralogues
- Sequences

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

How do I use BioMart?
4. Results

View and export data table/sequence in a number of formats:

- html
- csv
- tsv
- xls
- fasta

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

biomaRt: Bioconductor package for BioMart

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data using the R statistical programming language

- Easy installation:

```
source("http://bioconductor.org/biocLite.R")
biocLite("biomaRt")
```

Data Retrieval
Retrieving Data From ENSEMBL

BioMart for Data Mining

More information

Kinsella *et al.*

Ensembl BioMarts: a hub for data retrieval across taxonomic space

Database: the Journal of Biological Databases and Curation (2011)

europepmc.org/abstract/MED/21785142

Smedley *et al.*

BioMart – biological queries made easy

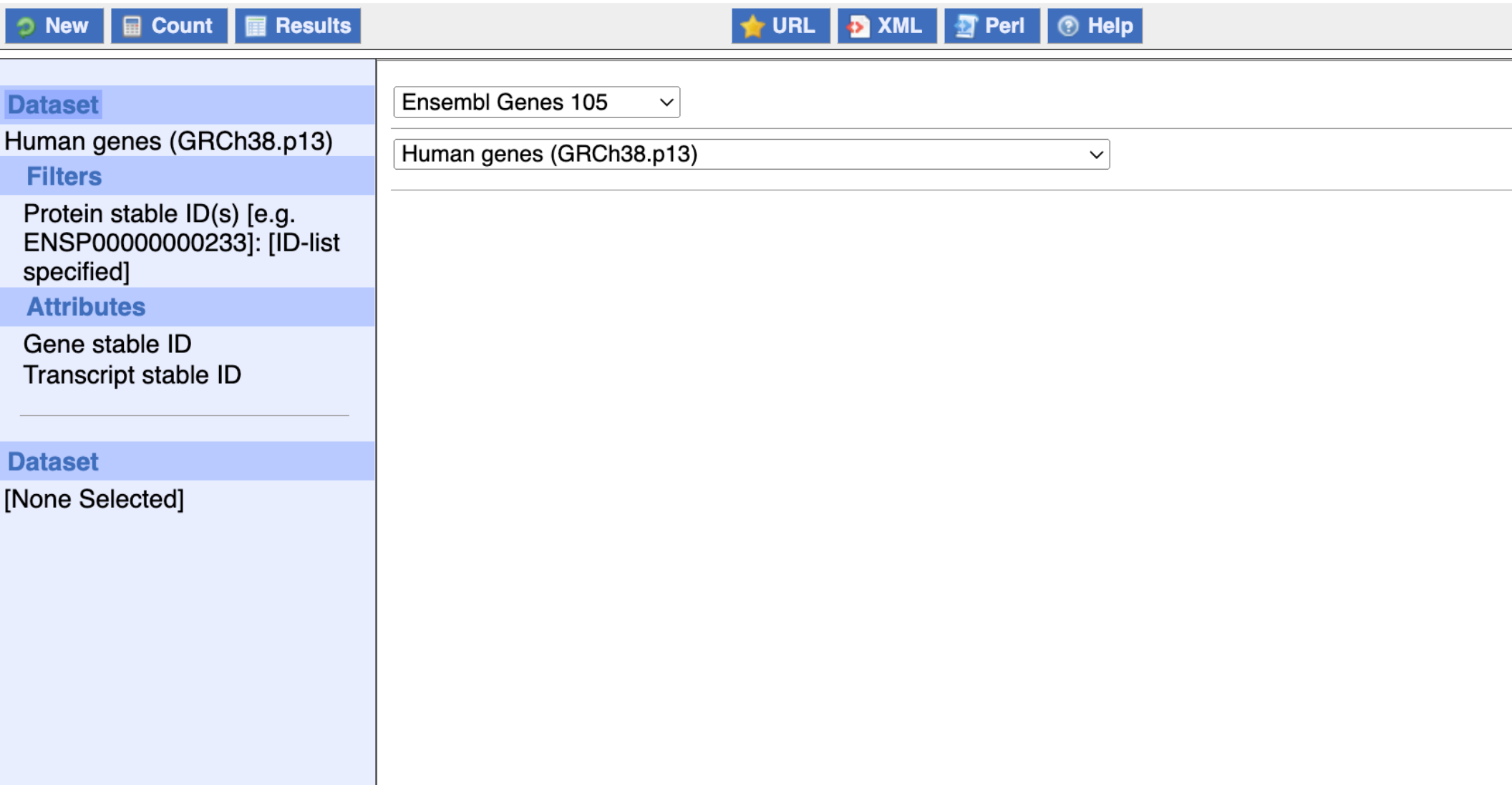
BMC Genomics (2009) 10:22

europepmc.org/abstract/MED/19144180

Data Retrieval
Retrieving Data From ENSEMBL

BioMart

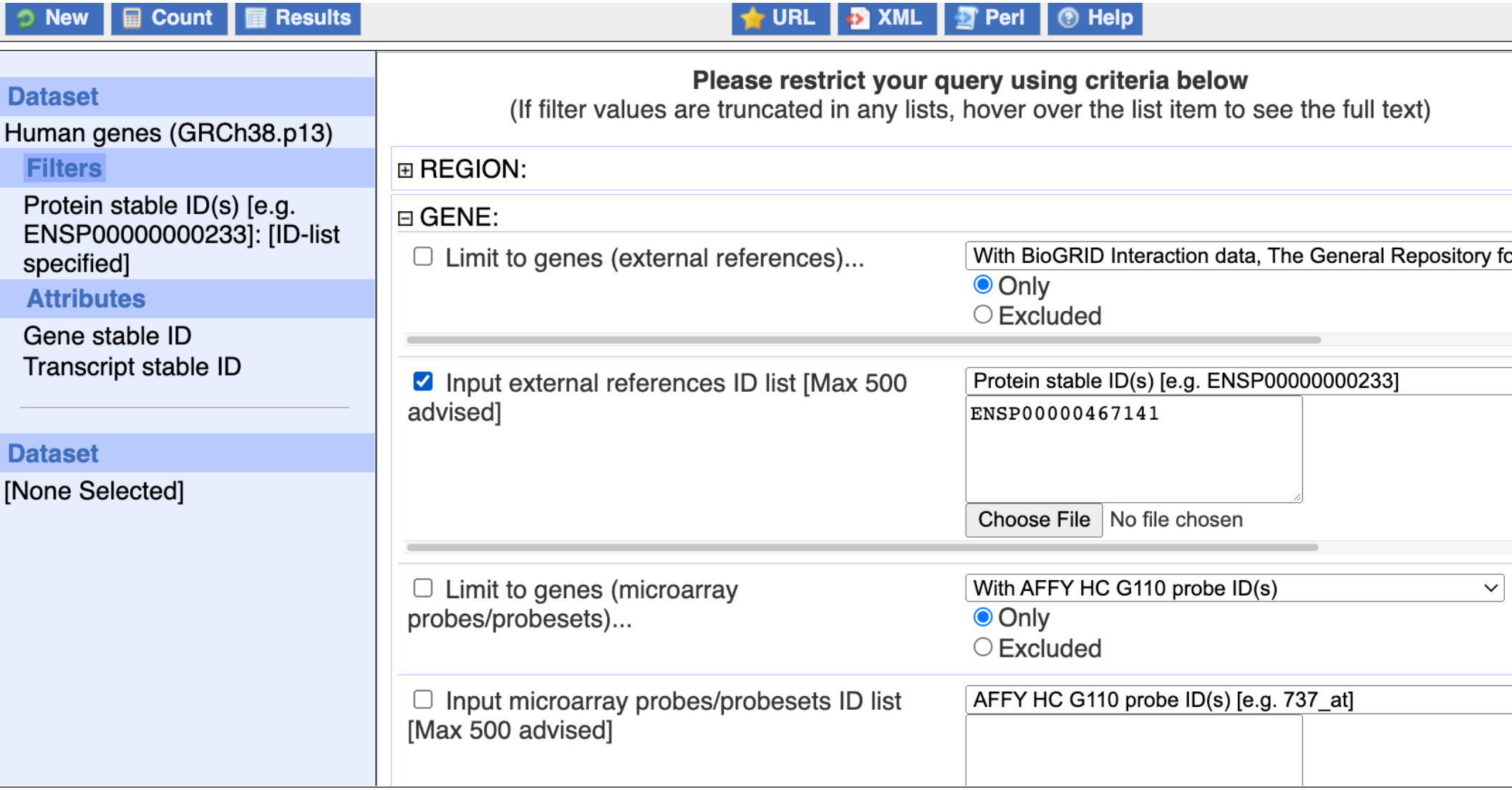
- Go to BioMart and Retrieve: **Human: ENSP00000467141**



Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- Go to BioMart and Retrieve: **Human: ENSP00000467141**



Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- Select: **Peptide**

New | Count | Results | URL | XML | Perl | Help

Dataset 1 / 68005 Genes
Human genes (GRCh38.p13)

Filters

Protein stable ID(s) [e.g. ENSP00000000233]: [ID-list specified]

Attributes

Peptide
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version

Dataset

[None Selected]

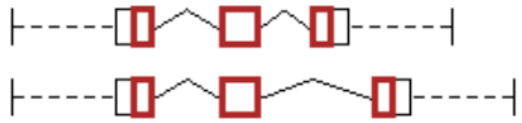
Please select columns to be included in the output and hit 'Results' when ready

Missing non coding genes in your mart query output, please check the following FAQ

- ○ **Features**
- ○ **Structures**
- ○ **Homologues (Max select 6 orthologues)**
- ○ **Variant (Germline)**
- ◉ **Sequences**

⊟ 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
- ◉ Peptide

Upstream flank

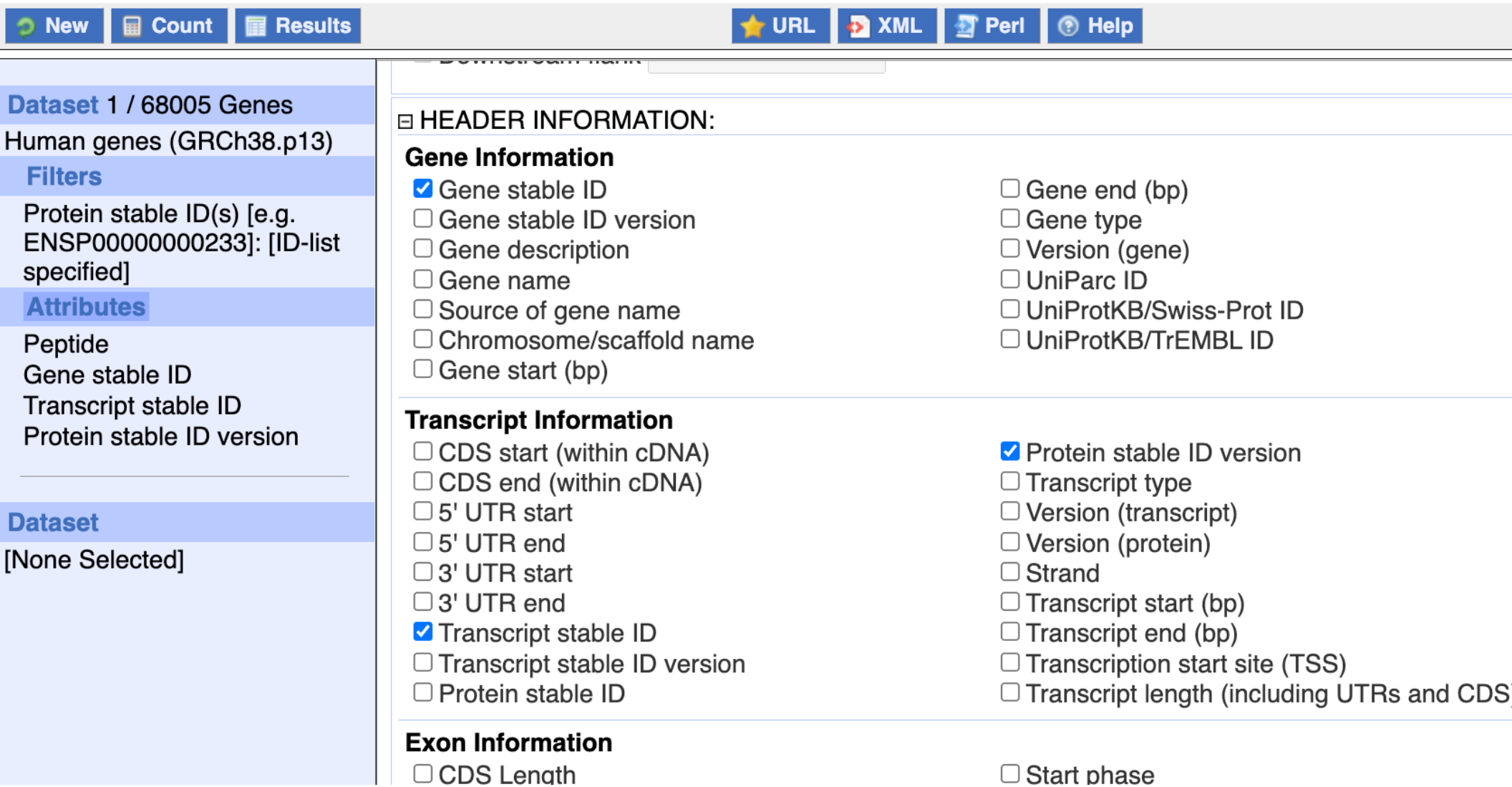
☐ Upstream flank

Downstream flank

Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- Select: **Gene stable ID, Transcript stable ID and Protein stable ID**



Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- Rename Download as: **Hs_ENSP00000467141.pep.fa**

New | Count | Results | URL | XML | Perl | Help

Dataset 1 / 68005 Genes
Human genes (GRCh38.p13)

Filters

Protein stable ID(s) [e.g. ENSP00000000233]: [ID-list specified]

Attributes

Peptide
Gene stable ID
Transcript stable ID
Protein stable ID version

Dataset

[None Selected]

Export all results to File FASTA Unique results only
Go

Email notification to

View 10 rows as FASTA Unique results only

```
>ENSG00000155657|ENST00000589042|ENSP00000467141.1
MTTQAPTFTQPLQSVVVLEGSTATFEAHISGFPVPEVSWFRDGQVISTSTLPGVQISFSD
GRAKLTIPAVTKANSGRYSLKATNGSGQATSTAELLVKAETAPPNFVQRLQSMTVRQGSQ
VRLQVRVTGIPTPVVKFYRDGAEIQSSLDFQISQEGDLYSLLIAEAYPEDSGTYSVNATN
SVGRATSTAELLVQGEEEVPAKKTKTIVSTAQISESRQTRIEKKIEAHFDARSIATVEMV
IDGAAGQQLPHKTPPRIPPKPKSRSPTPPSIAAKAQLARQQSPSPIRHSPSPVRHVRAPT
PSPVRSVSPAARISTSPIRSVRSPLLMRKTQASTVATGPEVPPPWKQEGYVASSSEAEMR
ETTLTTSTQIRTEERWEGRYGVQEQVTISGAAGAAASVSASASYAAEAVATGAKEVKQDA
DKSAAVATVVAAVDMARVREPVISAVEQTAQRTTTTAVHIQPAQEQVRKEAEKTAVTKVV
VAADKAKEQELKSRTKEVITTKQEQMHVTHEQIRKETEKTFVPKVVISAAKAKEQETRIS
EEITKKQKQVTQEAIRQETEITAASMVVVATAKSTKLETVPGAQEETTTQQDQMHLSYEK
IMKETRKTVVPKVIVATPKVKEQDLVSRGREGITTKREQVQITQEKMRKEAEKTALSTIA
VATAKAKEQETILRTRETMATRQEQIQVTHGKVDVGKKAEAVATVVAAVDQARVREPREP
GHLEESYAQQTTLEYGYKERISAAKVAEPPQRPASEPHVVPKAVKPRVIQAPSETHIKTT
DQKGMHISSQIKKTTDLTTERLVHVDKRPRTASPHFTVSKISVPKTEHGYEASIAGSAIA
TLQKELSATSSAQKITKSVKAPTVKPSETRVRAEPTPLPQFPFADTPDTYKSEAGVEVKK
EVGVSITGTTVREERFEVLHGREAKVTETARVPAPVEIPVTPPTLVSGLKNVTVIEGESV
TLECHISGYPSPTVTWYREDYQIESSIDFQITFQSGIARLMIREAFAEDSGRFTCSAVNE
AGTVSTSCYLAVQVSEEFEKETTAVTEKFTTEEKRFVESRDVVMTDTSLTEEQAGPGEPA
APYFITKPVVQKLVEGGSVVFGCQVGGNPKPHVYWKKSGVPLTTGYRYKVSYNKQTGECK
LVISMTFADDAGEYTIVVRNKHGETSASASLLEEADYELLMKSQQEMLYQTQVTAFVQEP
```

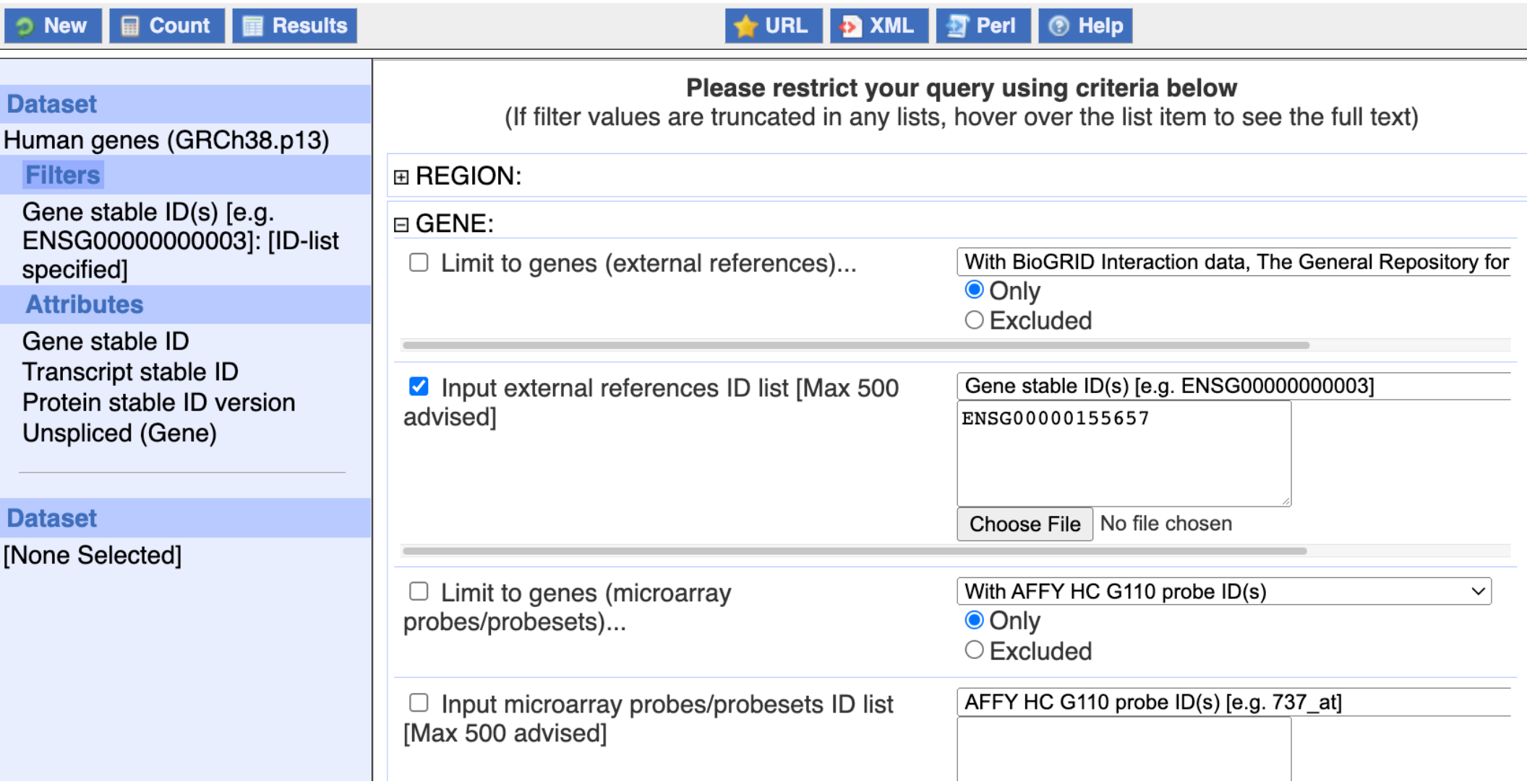
Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- From the Header we extract the following table of relationships:

  >ENSG00000155657|ENST00000589042|ENSP00000467141.1

- Now, we will proceed to extract: **ENSG00000155657**



Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- From the Header we extract the following table of relationships:

  >ENSG00000155657|ENST00000589042|ENSP00000467141.1

- Now, we will proceed to extract the: **Unspliced (Gene)**

New | Count | Results | URL | XML | Perl | Help

Dataset

Human genes (GRCh38.p13)

Filters

Protein stable ID(s) [e.g. ENSP00000000233]: [ID-list specified]

Attributes

Gene stable ID
Transcript stable ID
Protein stable ID version
Unspliced (Gene)

Dataset

[None Selected]

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

Missing non coding genes in your mart query output, please check the following FAQ

- ○ **Features**
- ○ **Structures**
- ○ **Homologues (Max select 6 orthologues)**
- ○ **Variant (Germline)**
- ◉ **Sequences**

⊟ 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
- ○ Peptide

Upstream flank

☐ Upstream flank

Downstream flank

Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- From the Header we extract the following table of relationships:

  >ENSG00000155657|ENST00000589042|ENSP00000467141.1

- Select: **Gene stable ID, Transcript stable ID and Protein stable ID**

New | Count | Results | URL | XML | Perl | Help

Dataset

Human genes (GRCh38.p13)

Filters

Protein stable ID(s) [e.g. ENSP00000000233]: [ID-list specified]

Attributes

Gene stable ID
Transcript stable ID
Protein stable ID version
Unspliced (Gene)

Dataset

[None Selected]

HEADER INFORMATION:

Gene Information

- [x] Gene stable ID
- [ ] Gene stable ID version
- [ ] Gene description
- [ ] Gene name
- [ ] Source of gene name
- [ ] Chromosome/scaffold name
- [ ] Gene start (bp)
- [ ] Gene end (bp)
- [ ] Gene type
- [ ] Version (gene)
- [ ] UniParc ID
- [ ] UniProtKB/Swiss-Prot ID
- [ ] UniProtKB/TrEMBL ID

Transcript Information

- [ ] CDS start (within cDNA)
- [ ] CDS end (within cDNA)
- [ ] 5' UTR start
- [ ] 5' UTR end
- [ ] 3' UTR start
- [ ] 3' UTR end
- [x] Transcript stable ID
- [ ] Transcript stable ID version
- [ ] Protein stable ID
- [x] Protein stable ID version
- [ ] Transcript type
- [ ] Version (transcript)
- [ ] Version (protein)
- [ ] Strand
- [ ] Transcript start (bp)
- [ ] Transcript end (bp)
- [ ] Transcription start site (TSS)
- [ ] Transcript length (including UTRs and CDS)

Exon Information

- [ ] CDS Length
- [ ] CDS start
- [ ] Start phase
- [ ] End phase

Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- Rename Download as: **Hs_ENSG00000155657.dna.fa**

New | Count | Results | URL | XML | Perl | Help

Dataset

Human genes (GRCh38.p13)

Filters

Gene stable ID(s) [e.g. ENSG00000000003]: [ID-list specified]

Attributes

Gene stable ID
Transcript stable ID
Protein stable ID version
Unspliced (Gene)

Dataset

[None Selected]

Export all results to File FASTA Unique results only

Go

Email notification to

View 10 rows as FASTA Unique results only

```
>ENSG00000155657|ENST00000589042;ENST00000591111;ENST00000342175;ENST000003
GTTCCAGTTCTGCTGAGACACAACTTCCCCTGGGAAGCCTCCTGACCCATCAAGTCCAGA
GTAGATGGTCCTGCCATGTGCACTGTCAGCACCTGCACTGCCCAGCCACGCCCCTGAATA
TACACAAAGGAATCGCAGGTTTGAGCATCAGTCTCCTCTGGCTAGGCTGTAAGAACCTTA
TCAGCAGGATTCATGACAGCCCTGTTGTCATTGTATTGCCAGTACCTAAGACACAGTCAC
ACTCAGTAAACTCTTTTGAATTGAATTGTTGCCAAAGGTAAACAAAGATAAACCTAGCTT
GTTTTTCTCCCCTTTTTAGAGGGAGGAGAGGGGAAATGTAAAAAGACAAAGATGATTCTC
CCTTCTTATAGACAGGAGTCAATCCGAATCAGGCCCTAACAATGAAGCACTCATGAAAAT
CTCTCACCTCATCTGCAAAACTTGTGTGAGAACCCTCCAGACCCTCTGGGCTAGACCCTG
ATTGAAAATTTGAATTTAAAAATAGATGCTGGCTCCACGGACTAAAATAATAGTAAGGAG
GAAAGAGGAAAGGTCCAGGATTGTTGTTCCTGGATCAATATTGTTCCAACATTTTAAAGT
GTTCCAACATTTTAAAACATATGTACTTAGGCATAAATCATCATCCGGTATGAGGAACTG
GATGCCTCCACTCTCTATGACTAAATCCAAACTGTCAAAACCAGAATTTATGTACTTCTT
TAGAACACTACGTAAATGATCAATATTTAATTGTAGCTGGGGCTCAGTAGTCTGGGCTGT
GAACAGATCTTATTTGAAACCAGTATTTCATACAAACTTGTAAATATAAAGAAAGCCAAA
TCCAAGAACTTAAGAGTATTTGAATTCCAAATGAAGAAAGGAAAGAATATCTTATTTTGT
TCTGCTTTATGTTTGTCTTTATTCATGACCATTTCTTACTAAAAATCACCATCTGAAGCT
TTTGAAACTCCTGTTGTATGCACAGCATTTTAAGTGCAGCCATGAAATAAGTTAGTTGAA
ACAAAATAAAATATCTCCTTCCAGTATGCTCTGATATTGTGTCAATGTAAATGCCAAGGC
ATGACAGACTGTTTTCCATTTATAATAAGCGGAATCTTGAGTTGCATTGGAAAGCTAAAA
GATTTGCGTTGACTTAAATGAAAGTGTCCTTTTTAAAGCAAAAACAATTCATTTTCTCTG
```

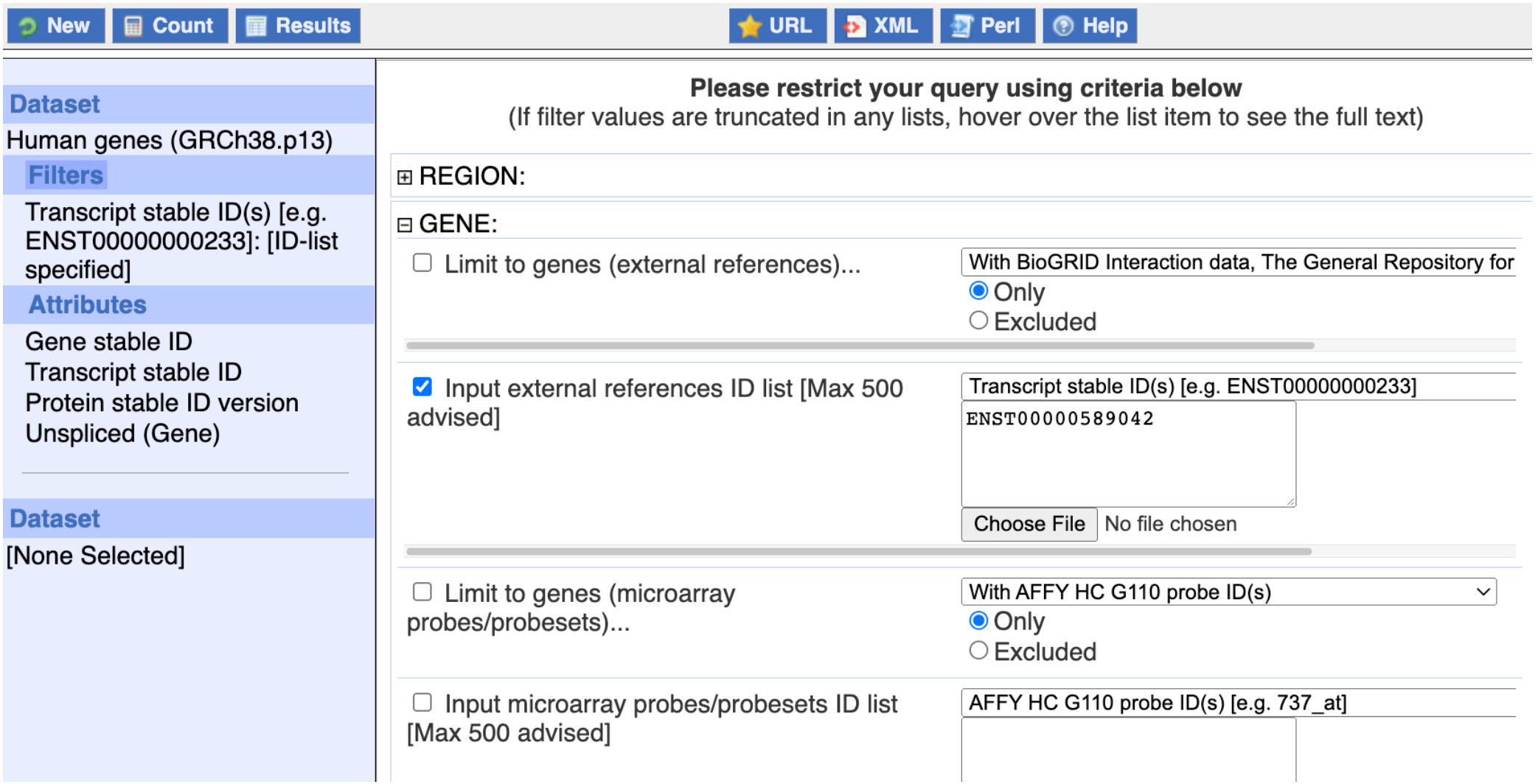
Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- From the Header we extract the following table of relationships:

  >ENSG00000155657|ENST00000589042|ENSP00000467141.1

- Now, we will proceed to extract: **ENST00000589042**



Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- From the Header we extract the following table of relationships:

  >ENSG00000155657|ENST00000589042|ENSP00000467141.1

- Now, we will proceed to extract the: **cDNA sequences**

New | Count | Results | URL | XML | Perl | Help

Dataset
Human genes (GRCh38.p13)

Filters
Transcript stable ID(s) [e.g. ENST00000000233]: [ID-list specified]

Attributes
Gene stable ID
Transcript stable ID
Protein stable ID version
cDNA sequences

Dataset
[None Selected]

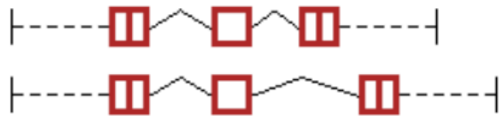
Please select columns to be included in the output and hit 'Results' when ready

Missing non coding genes in your mart query output, please check the following FAQ

- ○ **Features**
- ○ **Structures**
- ○ **Homologues (Max select 6 orthologues)**
- ○ **Variant (Germline)**
- ◉ **Sequences**

⊟ 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
- ○ Peptide

Upstream flank

☐ Upstream flank

Downstream flank

Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- From the Header we extract the following table of relationships:

  >ENSG00000155657|ENST00000589042|ENSP00000467141.1

- Now, we will proceed to extract the: **cDNA sequences**

New | Count | Results | URL | XML | Perl | Help

Dataset

Human genes (GRCh38.p13)

Filters

Transcript stable ID(s) [e.g. ENST00000000233]: [ID-list specified]

Attributes

Gene stable ID
Transcript stable ID
Protein stable ID version
cDNA sequences

Dataset

[None Selected]

HEADER INFORMATION:

Gene Information

- [x] Gene stable ID
- [ ] Gene stable ID version
- [ ] Gene description
- [ ] Gene name
- [ ] Source of gene name
- [ ] Chromosome/scaffold name
- [ ] Gene start (bp)
- [ ] Gene end (bp)
- [ ] Gene type
- [ ] Version (gene)
- [ ] UniParc ID
- [ ] UniProtKB/Swiss-Prot ID
- [ ] UniProtKB/TrEMBL ID

Transcript Information

- [ ] CDS start (within cDNA)
- [ ] CDS end (within cDNA)
- [ ] 5' UTR start
- [ ] 5' UTR end
- [ ] 3' UTR start
- [ ] 3' UTR end
- [x] Transcript stable ID
- [ ] Transcript stable ID version
- [ ] Protein stable ID
- [x] Protein stable ID version
- [ ] Transcript type
- [ ] Version (transcript)
- [ ] Version (protein)
- [ ] Strand
- [ ] Transcript start (bp)
- [ ] Transcript end (bp)
- [ ] Transcription start site (TSS)
- [ ] Transcript length (including UTRs and CDS)

Exon Information

- [ ] CDS Length
- [ ] Start phase

Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- Rename Download as: **Hs_ENST00000589042.rna.fa**

New | Count | Results | URL | XML | Perl | Help

Dataset
Human genes (GRCh38.p13)

Filters
Transcript stable ID(s) [e.g. ENST00000000233]: [ID-list specified]

Attributes
Gene stable ID
Transcript stable ID
Protein stable ID version
cDNA sequences

Dataset
[None Selected]

Export all results to File FASTA Unique results only
Go

Email notification to

View 10 rows as FASTA Unique results only

```
>ENSG00000155657|ENST00000589042|ENSP00000467141.1
GAGCAGTCGTGCATTCCCAGCCTCGCCTCGGGTGTAGGGATTGCATAGAAAAGCAAAACT
ACACAGTCTTGACTGTGTAGTTTTGTTTTTAGGATTAGAGGCTCACCGATTCATGTCGGA
GATGGTCAGAAAAACCAACTCTCCATAGGACGTCGTTTCAGAAGCAACCTTGGGCTTAGT
CCCACCCTTTTTAGGCACTCTTGAGAAATCAGAGTGCCTAGAAAGATGACAACTCAAGCA
CCGACGTTTACGCAGCCGTTACAAAGCGTTGTGGTACTGGAGGGTAGTACCGCAACCTTT
GAGGCTCACATTAGTGGTTTTCCAGTTCCTGAGGTGAGCTGGTTTAGGGATGGCCAGGTG
ATTTCCACTTCCACTCTGCCCGGCGTGCAGATCTCCTTTAGCGATGGCCGCGCTAAACTG
ACGATCCCCGCCGTGACTAAAGCCAACAGTGGACGATATTCCCTGAAAGCCACCAATGGA
TCTGGACAAGCGACTAGTACTGCTGAGCTTCTCGTGAAAGCTGAGACAGCACCACCCAAC
TTCGTTCAACGACTGCAGAGCATGACCGTGAGACAAGGAAGCCAAGTGAGACTCCAAGTG
AGAGTGACTGGAATCCCTACACCTGTGGTGAAGTTCTACCGGGATGGAGCCGAAATCCAG
AGCTCCCTTGATTTCCAAATTTCACAAGAAGGCGACCTCTACAGCTTACTGATTGCAGAA
GCATACCCTGAGGACTCAGGGACCTATTCAGTAAATGCCACCAATAGCGTTGGAAGAGCT
ACTTCGACTGCTGAATTACTGGTTCAAGGTGAAGAAGAAGTACCTGCTAAAAAGACAAAG
ACAATTGTTTCGACTGCTCAGATCTCAGAATCAAGACAAACCCGAATTGAAAAGAAGATT
GAAGCCCACTTTGATGCCAGATCAATTGCAACAGTTGAGATGGTCATAGATGGTGCCGCT
GGGCAACAGCTGCCACATAAAACACCTCCCAGGATTCCTCCGAAGCCAAAGTCAAGATCC
CCAACACCACCGTCTATTGCTGCCAAAGCACAGCTGGCTCGGCAGCAGTCCCCATCGCCC
ATAAGACACTCCCCTTCCCCGGTCAGACACGTGCGGGCACCGACCCCATCTCCGGTCAGG
TCCGTGTCTCCAGCAGCAAGAATCTCCACATCCCCCATCAGGTCTGTTAGGTCTCCATTG
```

Data Retrieval
Retrieving Data From ENSEMBL

BioMart

- Go to BioMart and Do the same for:

  Mouse: **ENSMUSP00000097561**

  - Name the resulting files as:
    - **Mn_ENSMUSP00000097561.pep.fa**
    - **Mm_ENSMUSG00000051747.dna.fa**
    - **Mn_ENSMUST00000099981.rna.fa**

Data Retrieval
Retrieving Data From Uniprot

- Go to: Uniprot
- Search for:

  danio rerio ribosomal protein 40s 60s
- In a different tab search for:

  danio rerio ribosomal protein 40s 60s AND reviewed:yes
- Download Results (Follow instructions)

  Fasta/Canonical/Compressed

  Fasta/Canonical & Isoform/Compressed
- Upload results into Kaiser
- Rename Datasets:

  DrRibosomalProtCan

  DrRibosomalProtCanIso
- Check File Type

  Go to ‘Edit Atributes’ > “Datatype’

  Make sure it is set to ‘fasta’

  Save