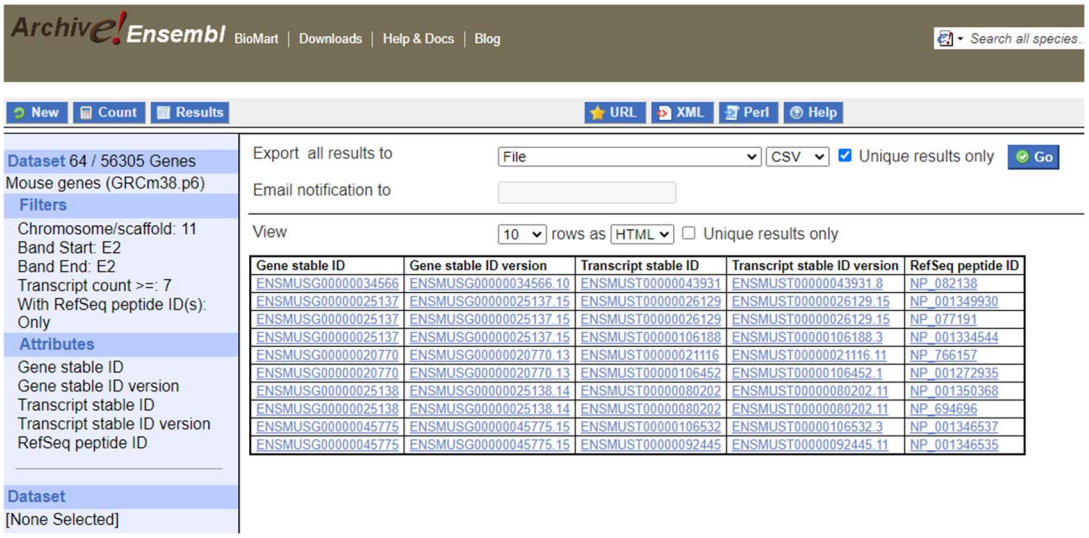
BIOMART

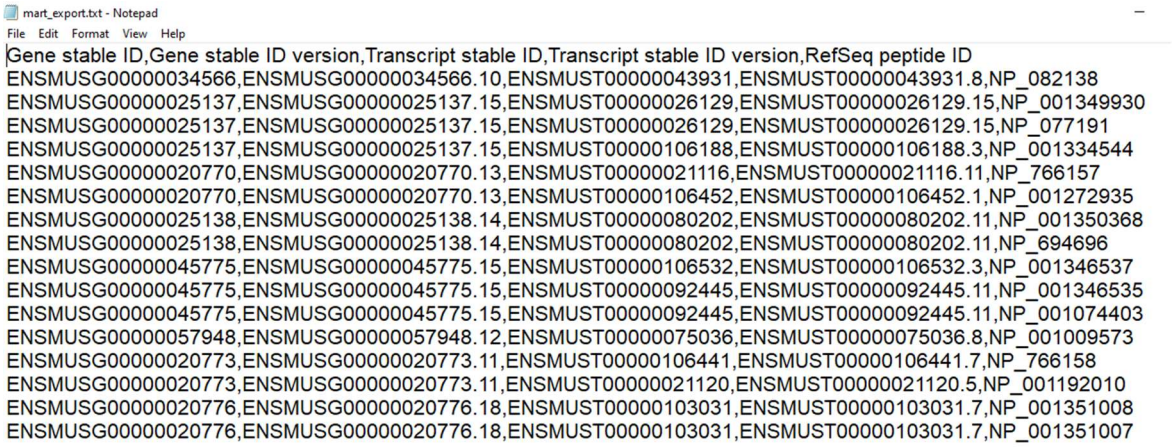
Use the web-based Biomart in Ensembl to create a dataset using the following parameters:

***Dataset*:**
Ensembl Genes 100 **(or the latest version)**
Mouse genes (GRCm38.p6) **(or the latest version)**
Filters:
Chromosome 11
Band E2 only
Transcript count >=7
Limit to genes with RefSeq protein (peptide) IDs only
Attributes:
Default attributes
Add "RefSeq Protein (peptide) ID"

I used the datasets Ensembl Genes 102 (from the archive site) and Mouse Genes (GRCm38.p6). Screenshots of the results are shown below



Gene stable ID	Gene stable ID version	Transcript stable ID	Transcript stable ID version	RefSeq peptide ID
ENSMUSG00000034566	ENSMUSG00000034566.10	ENSMUST00000043931	ENSMUST00000043931.8	NP_082138
ENSMUSG00000025137	ENSMUSG00000025137.15	ENSMUST00000026129	ENSMUST00000026129.15	NP_001349930
ENSMUSG00000025137	ENSMUSG00000025137.15	ENSMUST00000026129	ENSMUST00000026129.15	NP_077191
ENSMUSG00000025137	ENSMUSG00000025137.15	ENSMUST00000106188	ENSMUST00000106188.3	NP_001334544
ENSMUSG00000020770	ENSMUSG00000020770.13	ENSMUST00000021116	ENSMUST00000021116.11	NP_766157
ENSMUSG00000020770	ENSMUSG00000020770.13	ENSMUST00000106452	ENSMUST00000106452.1	NP_001272935
ENSMUSG00000025138	ENSMUSG00000025138.14	ENSMUST00000080202	ENSMUST00000080202.11	NP_001350368
ENSMUSG00000025138	ENSMUSG00000025138.14	ENSMUST00000080202	ENSMUST00000080202.11	NP_694696
ENSMUSG00000045775	ENSMUSG00000045775.15	ENSMUST00000106532	ENSMUST00000106532.3	NP_001346537
ENSMUSG00000045775	ENSMUSG00000045775.15	ENSMUST00000092445	ENSMUST00000092445.11	NP_001346535



BIOMART - R

Search OMIM.org for "huntington's disease". The first five entries all have this or a similar phrase in the title. Record the five identifiers (six-digit numbers) of those five records. The corresponding biomaRt filter name for these identifiers is "mim_morbid_accession". Use biomaRt to retrieve two tables with the following attributes, limiting to the five MIM values you found

First table
Entrez Gene ID
HGNC symbol
Ensembl Gene ID
Second table
HGNC symbol
Ensembl Gene ID
Ensembl Transcript ID

The top 5 identifiers are: 603218, 604802, 143100, 606438, 607136
The table sizes are different because the second one is showing transcripts for each gene, some of which have multiple transcripts

R Script

```
library(biomaRt)
ensembl = useMart("ENSEMBL_MART_ENSEMBL",dataset="hsapiens_gene_ensembl",
                  host="www.ensembl.org")
getBM(attributes= c("entrezgene_id", "hgnc_symbol",
      "ensembl_gene_id"), filters = "mim_morbid_accession",
      values=c("603218", "604802", "143100", "606438", "607136"),mart = ensembl)
getBM(attributes= c("hgnc_symbol", "ensembl_gene_id",
      "ensembl_transcript_id"), filters = "mim_morbid_accession",
      values=c("603218", "604802", "143100", "606438", "607136"),mart = ensembl)
```

Tables

```
> getBM(attributes= c("entrezgene_id", "hgnc_symbol",
+       "ensembl_gene_id"), filters = "mim_morbid_accession",
+       values=c("603218", "604802", "143100", "606438", "607136"),mart = ensembl)
  entrezgene_id hgnc_symbol ensembl_gene_id
1          3064         HTT ENSG00000197386
2          5621        PRNP ENSG00000171867
3         57338        JPH3 ENSG00000154118
4          6908         TBP ENSG00000112592
> getBM(attributes= c("hgnc_symbol", "ensembl_gene_id",
+       "ensembl_transcript_id"), filters = "mim_morbid_accession",
+       values=c("603218", "604802", "143100", "606438", "607136"),mart = ensembl)
   hgnc_symbol ensembl_gene_id ensembl_transcript_id
1          HTT ENSG00000197386       ENST00000680239
2          HTT ENSG00000197386       ENST00000680956
3          HTT ENSG00000197386       ENST00000680360
4          HTT ENSG00000197386       ENST00000681528
5          HTT ENSG00000197386       ENST00000647962
6          HTT ENSG00000197386       ENST00000649900
7          HTT ENSG00000197386       ENST00000680291
8          HTT ENSG00000197386       ENST00000355072
9          HTT ENSG00000197386       ENST00000648150
10         HTT ENSG00000197386       ENST00000506137
11         HTT ENSG00000197386       ENST00000512909
12         HTT ENSG00000197386       ENST00000510626
```