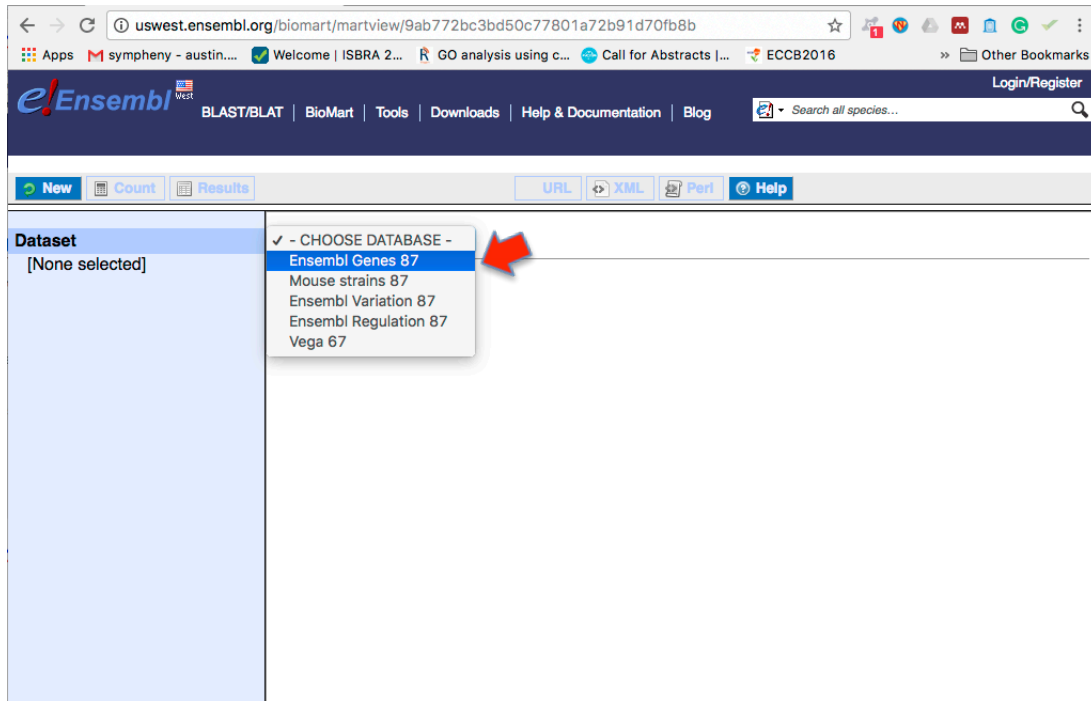
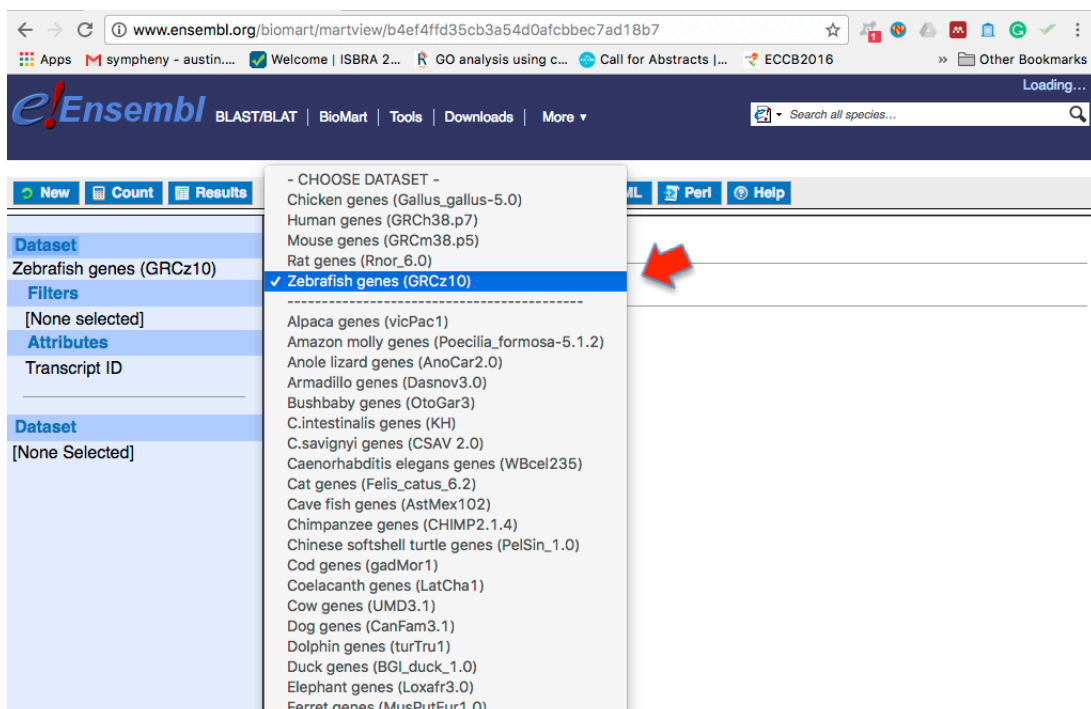


Tutorial on obtaining ID-mapping file

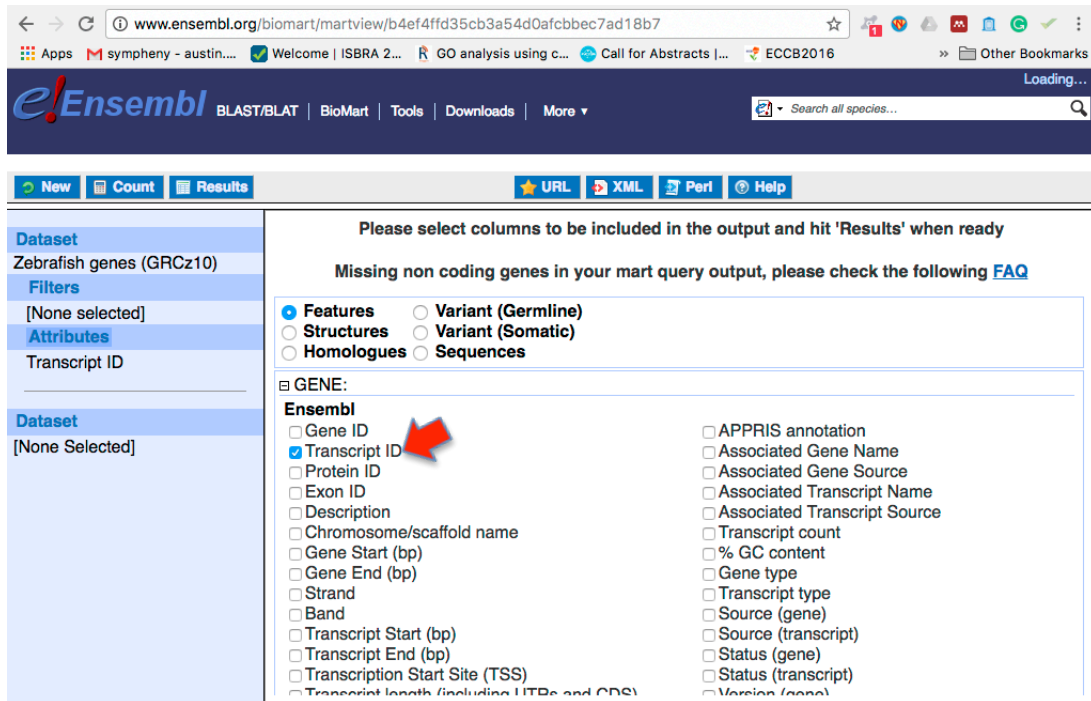
1. Go to the website of “**BioMart**”: <http://www.ensembl.org/biomart/martview/>
2. Choose the database: “**Ensembl Genes 87**”



3. Specify the dataset (i.e., species), for example: “**Zebrafish**”



4. Specify the attribute (i.e., ID type), for example: “**Transcript ID**”



www.ensembl.org/biomart/martview/b4ef4ffd35cb3a54d0afcbbec7ad18b7

Apps | symphony - austin... | Welcome | ISBRA 2... | GO analysis using c... | Call for Abstracts |... | ECCB2016 | Other Bookmarks

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | More ▾

Search all species...

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

Dataset
Zebrafish genes (GRCz10)

Filters
[None selected]

Attributes
Transcript ID

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](#)

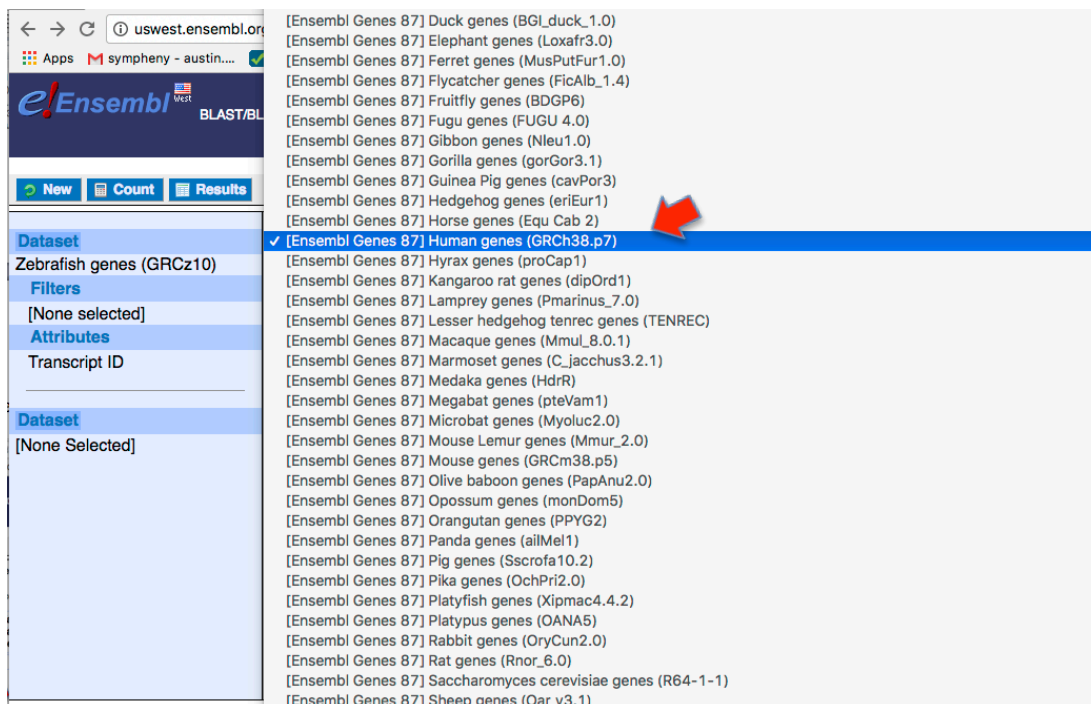
☒ **Features** ☐ **Variant (Germline)**
☐ **Structures** ☐ **Variant (Somatic)**
☐ **Homologues** ☐ **Sequences**

☐ **GENE:**

Ensembl

☐ Gene ID ☐ APPRIS annotation
☒ **Transcript ID** ☐ Associated Gene Name
☐ Protein ID ☐ Associated Gene Source
☐ Exon ID ☐ Associated Transcript Name
☐ Description ☐ Associated Transcript Source
☐ Chromosome/scaffold name ☐ Transcript count
☐ Gene Start (bp) ☐ % GC content
☐ Gene End (bp) ☐ Gene type
☐ Strand ☐ Transcript type
☐ Band ☐ Source (gene)
☐ Transcript Start (bp) ☐ Source (transcript)
☐ Transcript End (bp) ☐ Status (gene)
☐ Transcript Start Site (TSS) ☐ Status (transcript)
☐ Transcript length (including UTRs and CDS) ☐ Version (gene)

5. Specify the dataset (i.e., species) you want to convert, for example: “**Human**”



uswest.ensembl.org

Apps | symphony - austin... | Welcome | ISBRA 2... | GO analysis using c... | Call for Abstracts |... | ECCB2016 | Other Bookmarks

e!Ensembl BLAST/BLAT

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

Dataset
Zebrafish genes (GRCz10)

Filters
[None selected]

Attributes
Transcript ID

Dataset
[None Selected]

[Ensembl Genes 87] Duck genes (BGI_duck.1.0)
[Ensembl Genes 87] Elephant genes (Loxafr3.0)
[Ensembl Genes 87] Ferret genes (MusPutFur1.0)
[Ensembl Genes 87] Flycatcher genes (FicAlb.1.4)
[Ensembl Genes 87] Fruitfly genes (BDGP6)
[Ensembl Genes 87] Fugu genes (FUGU 4.0)
[Ensembl Genes 87] Gibbon genes (Nleu1.0)
[Ensembl Genes 87] Gorilla genes (gorGor3.1)
[Ensembl Genes 87] Guinea Pig genes (cavPor3)
[Ensembl Genes 87] Hedgehog genes (eriEur1)
[Ensembl Genes 87] Horse genes (Equ Cab 2)
☒ **[Ensembl Genes 87] Human genes (GRCh38.p7)**
[Ensembl Genes 87] Hyrax genes (proCap1)
[Ensembl Genes 87] Kangaroo rat genes (dipOrd1)
[Ensembl Genes 87] Lamprey genes (Pmarinus_7.0)
[Ensembl Genes 87] Lesser hedgehog tenrec genes (TENREC)
[Ensembl Genes 87] Macaque genes (Mmul_8.0.1)
[Ensembl Genes 87] Marmoset genes (C_jacchus3.2.1)
[Ensembl Genes 87] Medaka genes (HdrR)
[Ensembl Genes 87] Megabat genes (pteVam1)
[Ensembl Genes 87] Microbat genes (Myoluc2.0)
[Ensembl Genes 87] Mouse Lemur genes (Mmur_2.0)
[Ensembl Genes 87] Mouse genes (GRCm38.p5)
[Ensembl Genes 87] Olive baboon genes (PapAnu2.0)
[Ensembl Genes 87] Opossum genes (monDom5)
[Ensembl Genes 87] Orangutan genes (PPYG2)
[Ensembl Genes 87] Panda genes (ailMel1)
[Ensembl Genes 87] Pig genes (Sscrofa10.2)
[Ensembl Genes 87] Pika genes (OchPri2.0)
[Ensembl Genes 87] Platyfish genes (Xipmac4.4.2)
[Ensembl Genes 87] Platypus genes (OANA5)
[Ensembl Genes 87] Rabbit genes (OryCun2.0)
[Ensembl Genes 87] Rat genes (Rnor_6.0)
[Ensembl Genes 87] Saccharomyces cerevisiae genes (R64-1-1)
[Ensembl Genes 87] Sheep genes (Oar_v3.1)

6. Specify the attribute (i.e., ID type) as : “**Associated Gene Name**”

The screenshot shows the Ensembl BioMart interface. On the left, the 'Attributes' section for 'Human genes (GRCh38.p7)' is expanded, and 'Associated Gene Name' is selected. In the main area, under 'Please select columns to be included in the output', the 'GENE' section is expanded. Within the 'Ensembl' sub-section, 'Associated Gene Name' is checked, indicated by a red arrow. Other options like 'Transcript ID', 'Protein ID', and 'Exon ID' are also visible but not selected.

7. See if everything is OK by pressing the “**Results**” button:

The screenshot shows the Ensembl BioMart interface after clicking the 'Results' button. The 'Results' button is highlighted with a red arrow. Below it, there are options to export results (File, TSV) and email notifications. The 'View' section shows 10 rows as HTML. The resulting data table is as follows:

Transcript ID	Associated Gene Name
ENSDART00000093623	MT-ND6
ENSDART00000093600	MT-ND2
ENSDART00000120758	SNORD46
ENSDART00000093606	MT-CO1
ENSDART00000120228	SNORA68
ENSDART00000118593	SNORD35B
ENSDART00000118162	SNORA32
ENSDART00000119609	SNORA9
ENSDART00000093613	MT-CO3
ENSDART00000171547	TMEM39A

8. If everything looks good, export the results to the IDmapping file:
 - a. Please select the “TSV” format.
 - b. Please check the “Unique results only”.
 - c. Press “Go”.

The screenshot shows the Ensembl BioMart interface. On the left, the 'Dataset' is 'Zebrafish genes (GRCz10)' and the 'Attributes' are 'Transcript ID'. On the right, the 'Export all results to' dropdown is set to 'File', the format is 'TSV', and the 'Unique results only' checkbox is checked. Red arrows labeled 'a.', 'b.', and 'c.' point to the 'Unique results only' checkbox, the 'TSV' dropdown, and the 'Go' button respectively.

Transcript ID	Associated Gene Name
ENSDART00000093623	MT-ND6
ENSDART00000093600	MT-ND2
ENSDART00000120758	SNORD46
ENSDART00000093606	MT-CO1
ENSDART00000120228	SNORA68
ENSDART00000118593	SNORD35B
ENSDART00000118162	SNORA32
ENSDART00000119609	SNORA9
ENSDART00000093613	MT-CO3
ENSDART00000171547	TMEM39A

9. Rename the output file and column names:

- a. IDmapping file: [zebrafish2human_idmapping.txt](#)
- b. The column of your transcript as “ID” and the column for human gene name as “Human_genename”.

ID	Human_genename
ENSDART00000108990	PEX5L
ENSDART00000167199	RN7SL187P
ENSDART00000167277	RNA5SP406
ENSDART00000167199	RN7SL293P
ENSDART00000093618	MT-ND4
ENSDART00000165342	CES1
ENSDART00000164729	SBSPON
ENSDART00000167199	RN7SL87P
ENSDART00000146377	ZNF629

10. Assign this IDmapping file in your script (see the example files:

[Run_gsea_ex1.r](#) and [Run_gsea_ex2.r](#)):

```
#####  
## Setting Parameter  
#####  
#' @param DE_out_path the path of the DESeq2 output files  
#' @param export_path the path that user want to export the GSEA results  
#' @param convert 'TRUE': need to do id converion; 'FALSE': no id conversion needed  
#' @param IDmappingfile the idmapping file show that user's id to human gene name  
#' @param gene_name 'TRUE': there is a column "gene_name" in the input files (i.e. DESeq2 output files)  
#' @param GSEA_set = c('c0', 'c2', 'c3')  
#'          co - hallmark gene sets;      c1 - positional gene sets;  
#'          c2 - curated gene sets;      c3 - motif gene sets;  
#'          c4 - computational gene sets; c5 - GO gene sets;  
#'          c6 - oncogenic signatures;    c7 - immunologic signatures.  
#'          (http://software.broadinstitute.org/gsea/msigdb/collections.jsp#C2)  
#' @param minGSSize minimum size of genes for a specific GSEA gene set  
#' @param pAdjustMethod one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"  
#' @param pvalueCutoff cutoff of pvalue in the GSEA enrichment analysis  
#' @param export option for exporting GSEA result: "EXCEL", "PDF", or "BOTH"  
#' @param ShowTop 'TRUE': the visualization will display only the top and bottom N gene sets;  
#'          'FALSE': all of the significant gsea gene sets will be displayed  
#'          --> Note that: it is usually results in too many gene sets, which are not easy to see anything  
#' @param TopN this parameter indicate how many gene sets need to be displayed in the plots  
#####  
  
DE_out_path <- "Data/DE_out/"  
export_path <- "Results/DE_out_test/"  
IDmappingfile <- "Data/zebrafish2human_idmapping.txt"  
convert = FALSE  
gene_name = TRUE  
GSEA_set = c('c0', 'c2', 'c3')  
minGSSize = 30  
pAdjustMethod = "none"  
pvalueCutoff = 0.01  
export = "BOTH"  
ShowTop = TRUE  
TopN = 5
```