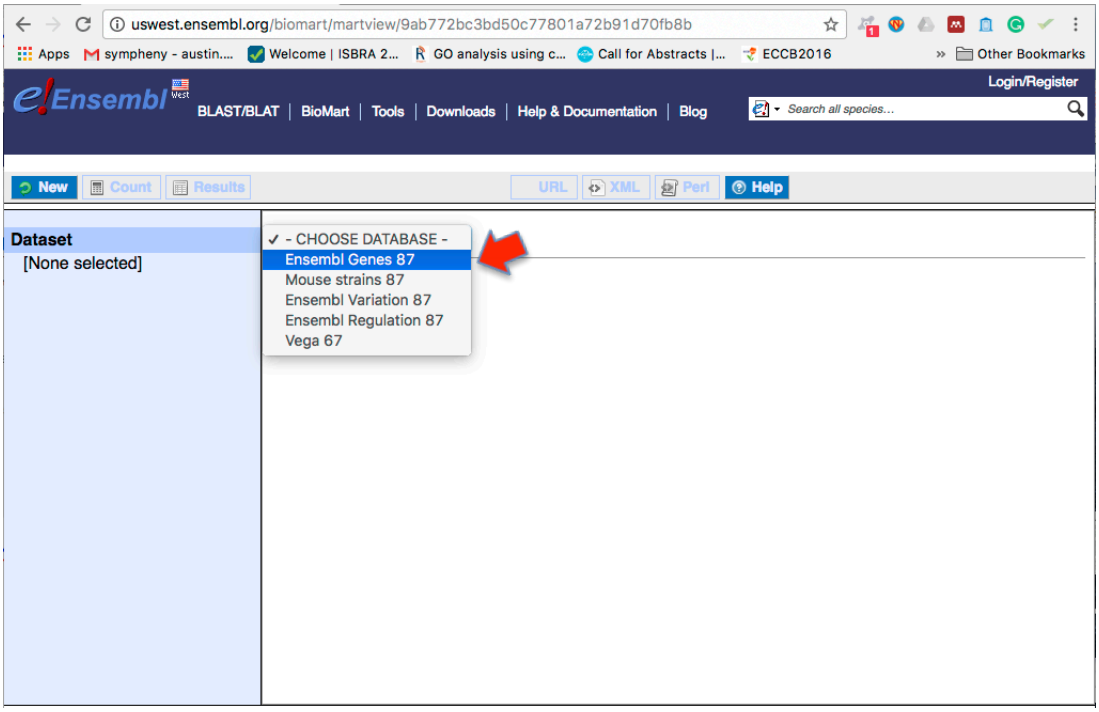
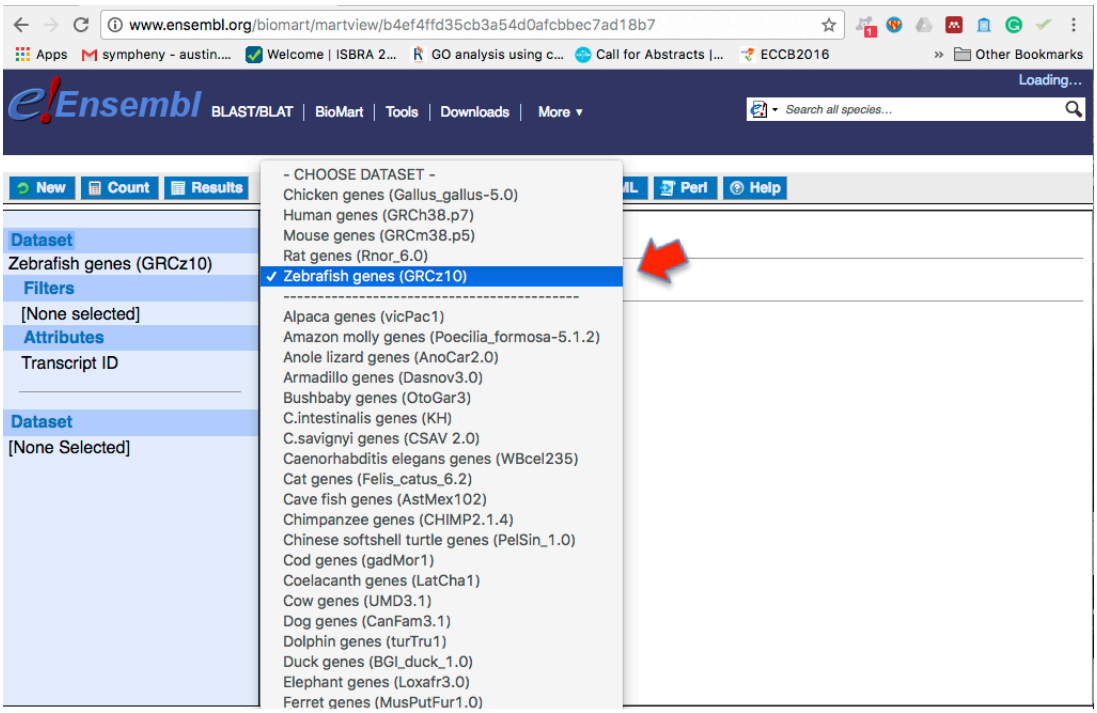
Tutorial on obtaining ID-mapping file

1. Go to the website of "**BioMart**": http://www.ensembl.org/biomart/martview/

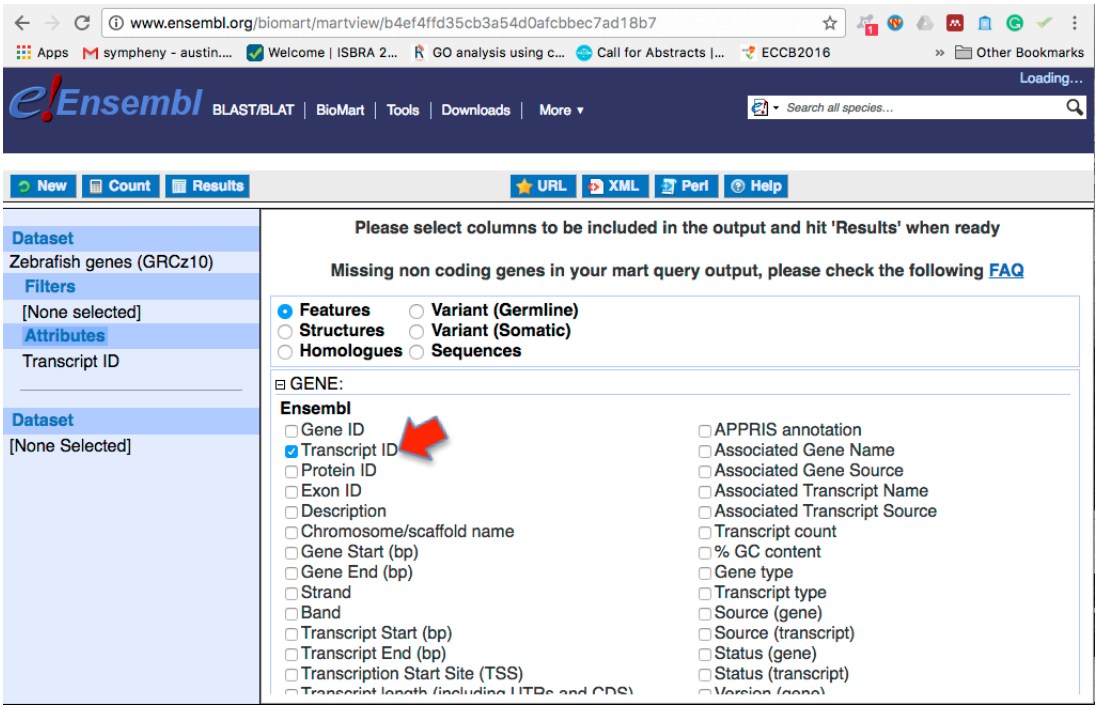
2. Choose the database: "**Ensembl Genes 87**"



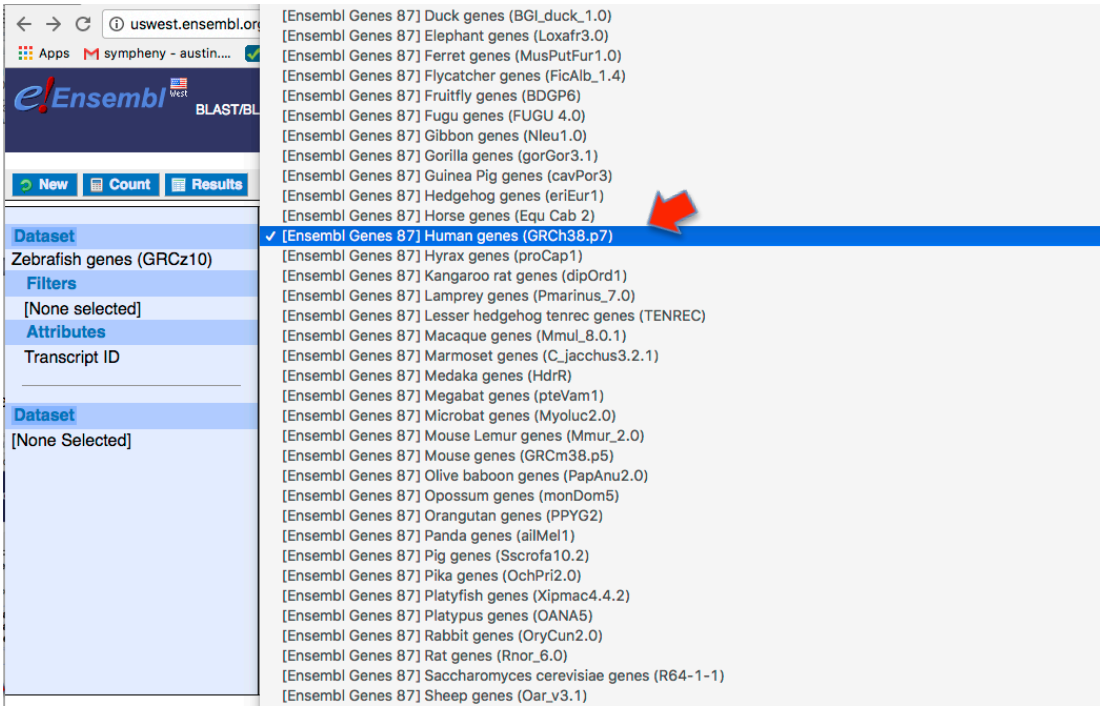
3. Specify the dataset (i.e., species), for example: "**Zebrafish**"



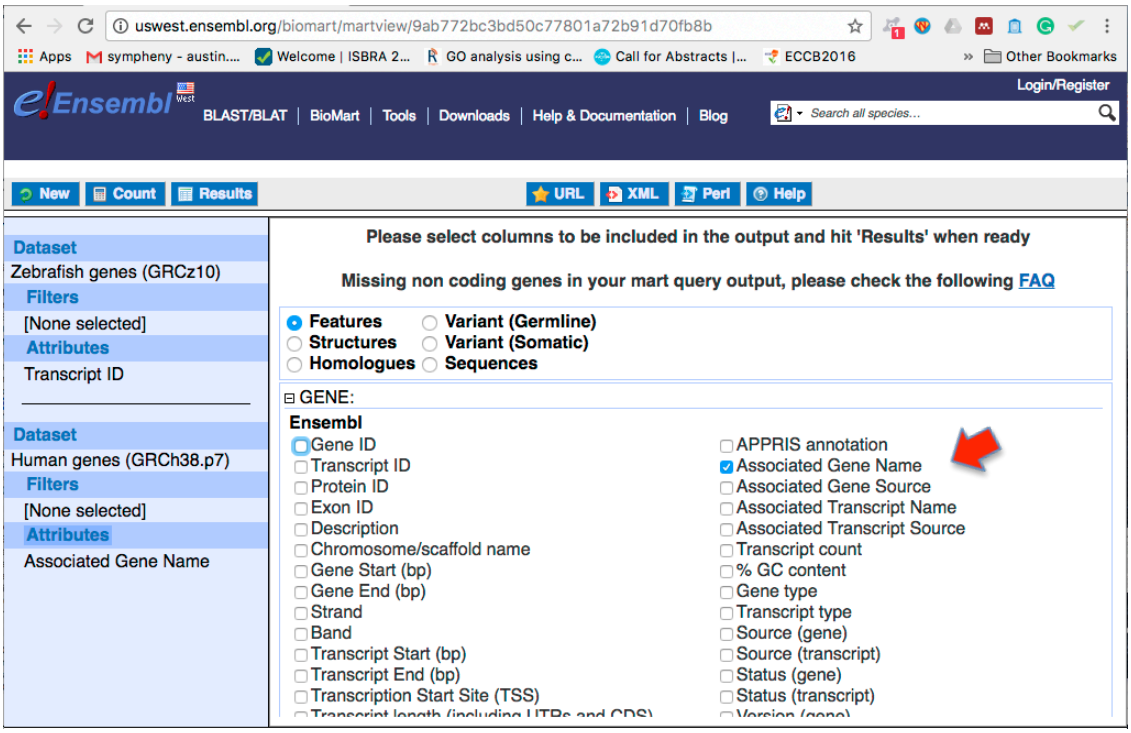
4. Specify the attribute (i.e., ID type), for example: "**Transcript ID**"



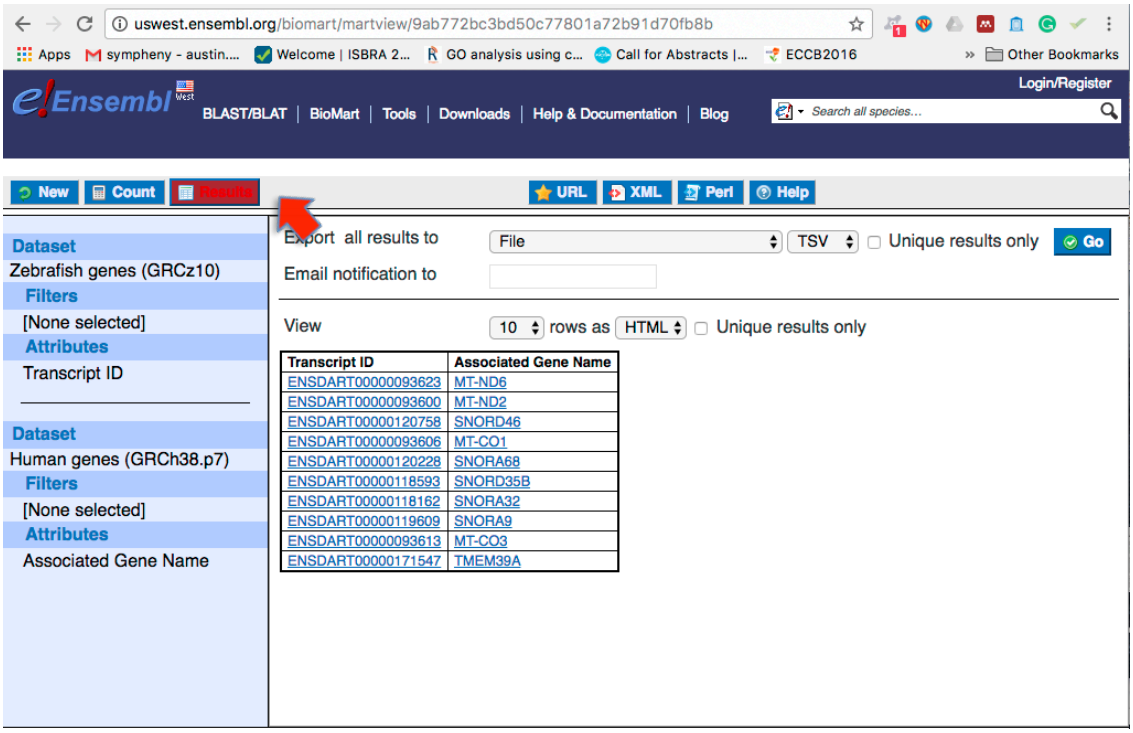
5. Specify the dataset (i.e., species) you want to convert, for example: "**Human**"



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



7. See if everything is OK by pressing the "**Results**" botton:

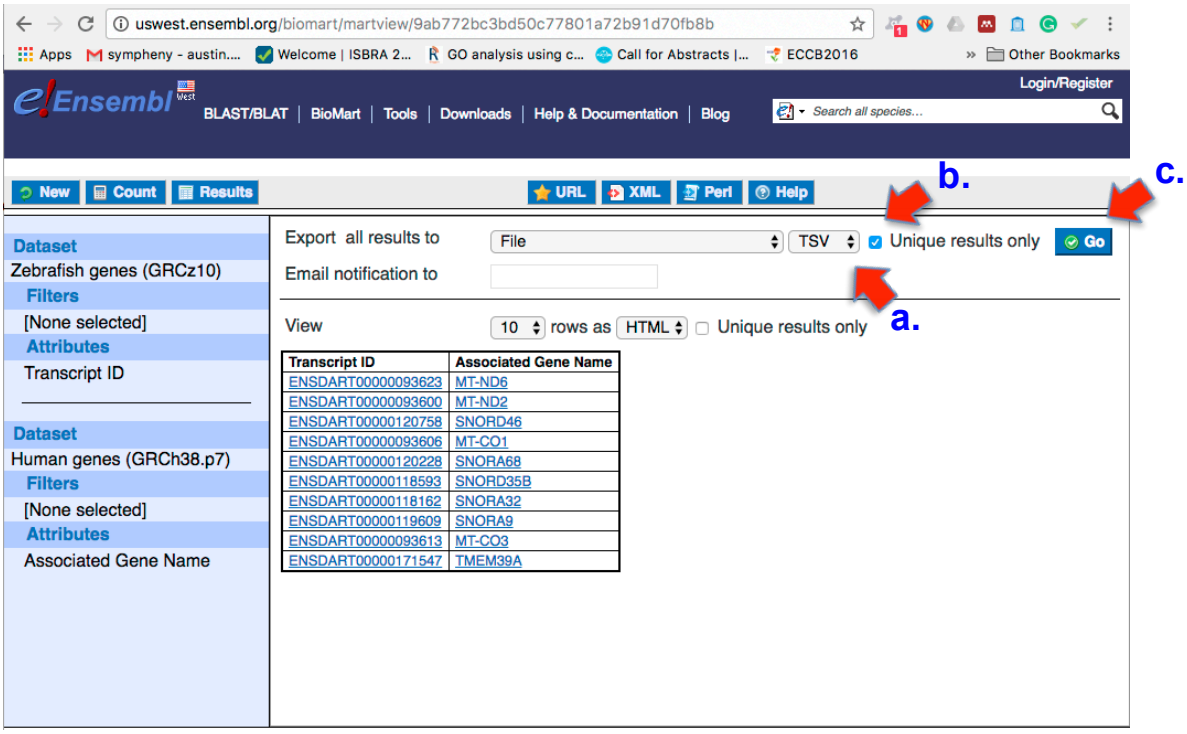


8. If everything looks good, export the results to be the IDmapping file:

    a. Please select the “TSV” format.

    b. Please check the “Unique results only”.

    c. Press “Go”.



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 columan "gene_name" in the input files (i.e. DESeq2 output files)
#' @param GSEA_set = c('c0', 'c2', 'c3')
#'                 c0 - 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
```