Doc for retrieving and preparing annotations for the RNA-Seq analysis

**Annotation columns in the Excel results file:**

Ensembl:

Download from Ensembl/BioMart for the relevant organism: Gene stable ID, Gene name, Gene type, Gene description.

Store as file called, for example, “mart\_export.txt”

Currently the file is placed in the project directory and the file name is defined in the global\_parameters.Rmd script

It is also possible to use the file “Annotation.tab” created by the updated “Retrieve\_annotation.Rmd” script, but this has not been tested yet

Trinotate:

Copy the file produced by trinotate ….trino\_anno\_rep.xls to the project directory. The file name is defined in the global\_parameters.Rmd script

--

**Enrichment analysis - KEGG**

kegg data are taken from Vered Perl script.

The perl script is located on Vered’s PC, at: Programming/Perl\_workspace/Assaf\_Rudich\_03\_HFD\_KEGG\_from\_Vered\_script/KEGG\_2\_ensembl\_conversion8.pl

[A newer script is mentioned in the most updated presentation of Assaf’s project:

create\_colored\_KEGG\_maps\_and\_summarize\_per\_path7.pl

(but this script is doing one step further, and uses the results of the previous script, therefore it is not relevant here)]

I copy the script KEGG\_2\_ensembl\_conversion8.pl to a new directory:



And rename it to “Retrieve\_KEGG\_annotation.pl”

**Enrichment analysis – GO**

GO data and the Annotation.tab file are from the R script.

The most updated R script I found was here:



05.Retrieve\_ensembl\_and\_kegg\_annot.Rmd

I copy it to the git directory and rename it to:

Retrieve\_annotation.Rmd

I did a first commit with the file as is, and them cleaned and tested it and did a second commit