**IMPORTANT NOTE:**

When analysis codes are integrated into the server, paths have to be updated in each R script file.

These variables are called datadir, outdir, setwd, targetdir etc. You would see them in the beginning of the file usually.

**LOGGING**

Every analysis code takes as input a job id parameter and creates two output log files:

job\_id\_analysis0\_info.txt: this file stores the output of regular print messages (includes messages such as “this gene is not found in the expression data etc.”)

job\_id\_analysis0\_error.txt: this file stores error messages given by R

**RESULTS**

You can see example results files in ‘results’ folder of each analysis folder. Check the most recent ones.

**ANALYSIS 0**

Description: plots the expression of the lncRNA or the RBP (

/usr/bin/Rscript analysis0\_plot/analysis0.R [gene\_id] [dataset\_id] [job\_id]

or if user enters target and background gene lists in the input page [optional]

/usr/bin/Rscript analysis0\_plot/analysis0.R [gene\_id] [dataset\_id] [job\_id] [target\_file] [background\_file]

as dataset\_id, we can enter

‘EMTAB2706’

‘EMTAB2770’

or

GTEX tissue specific files: (I removed the tissues that have less than 50 samples )

Sample counts can be found in Table3\_expression\_datasets/GTEX\_tissue\_counts\_samples.txt’de.. the ones with \* sign are too small and filtered out.

This is the list that we should work with:

GTEX\_Adipose\_Tissue.RData

GTEX\_Adrenal\_Gland.RData

GTEX\_Blood.RData

GTEX\_Blood\_Vessel.RData

GTEX\_Brain.RData

GTEX\_Breast.RData

GTEX\_Colon.RData

GTEX\_Esophagus.RData

GTEX\_Heart.RData

GTEX\_Liver.RData

GTEX\_Lung.RData

GTEX\_Muscle.RData

GTEX\_Nerve.RData

GTEX\_Ovary.RData

GTEX\_Pancreas.RData

GTEX\_Pituitary.RData

GTEX\_Prostate.RData

GTEX\_Salivary\_Gland.RData

GTEX\_Skin.RData

GTEX\_Small\_Intestine.RData

GTEX\_Spleen.RData

GTEX\_Stomach.RData

GTEX\_Testis.RData

GTEX\_Thyroid.RData

GTEX\_Uterus.RData

GTEX\_Vagina.RData

**Note: How do we determine gene\_id and dataset\_id?**

**We will have two tabs in analysis pop-up window for this analysis. In the first tab, we can plot the expression of the lncRNA, in the second tab we can plot the expression of the RBP.**

**RBP id is chosen as input by the user from the pulldown menu, just use that.**

**lncRNA id depends on from which row analysis button is clicked in the results page.**

**dataset\_id is also chosen by the user when s/he clicks on the analysis button, before showing anything we need to ask the user which dataset s/he is interested in**

**There should be three options:**

**EMTAB2706**

**EMTAB2770**

**GTEX**

**GTEX should have a subselection option where a tissue needs to be selected, this could be a pulldown menu with tissue name and size of that tissue (this info is available in** Table3\_expression\_datasets/GTEX\_tissue\_counts\_samples.txt (ignore the lines with \*) **)**

**Once the user selects one of these options we can pass that as argument to analysis0.R**

**for GTEX we need to pass it like this:**

**GTEX\_[tissue\_name] e.g. GTEX\_Liver**

**Note:** Because EMTAB2770 file has ~160 columns, I divided the plot into two:

see analysis0\_plot/results/ ENSG00000055917\_analysis0\_EMTAB2770\_part1.jpeg and ENSG00000055917\_analysis0\_EMTAB2770\_part2.jpeg

So, when we’re displaying the figures we need to be careful only for EMTAB2770, we need to display two figures at the same time.

Note: NO TXT FILE

**ANALYSIS 1 –correlation analysis**

/usr/bin/Rscript analysis1\_correlation/analysis1.R [lncRNA\_id] [RBP\_id] [dataset\_id] [job\_id] or

/usr/bin/Rscript analysis1\_correlation/analysis1.R [lncRNA\_id] [RBP\_id] [dataset\_id] [job\_id] [target\_file] [background\_file]

Output a figure + txt file

display the figure, give the option to download the txt file.

**ANALYSIS 2 –regression analysis**

/usr/bin/Rscript analysis2\_regression/analysis2.R [lncRNA\_id] [RBP\_id] [dataset\_id] [job\_id] or

/usr/bin/Rscript analysis2\_regression/analysis2.R [lncRNA\_id] [RBP\_id] [dataset\_id] [job\_id] [target\_file] [background\_file]

Output a figure + txt file

display the figure, give the option to download the txt file.

**ANALYSIS 3 –only for those lncRNAs with knockdown datasets [These are listed in Table5\_lncRNA\_KD\_datasets/ENSG\_ids\_with\_KD\_data.txt]**

/usr/bin/Rscript analysis3\_lncRNA\_KD\_CDF\_analysis/analysis3.R [lncRNA\_id] [RBP\_id] [job\_id] or

/usr/bin/Rscript analysis3\_lncRNA\_KD\_CDF\_analysis/analysis3.R [lncRNA\_id] [RBP\_id] [job\_id][target\_file] [background\_file]

Output a figure + txt file

display the figure, give the option to download the txt file.

NOTE: Some lncRNAs have more than one knockdown data, in that case there would be two figures and two txt files output. The corresponding tab should be arranged accordingly.

**R INSTALLATION INSTRUCTIONS**

The packages that I use:

**data.table**

**ggplot2**

**DAAG**

You should install DAAG from the code below, as otherwise it gives an error:

# devtools::install\_github('gokceneraslan/DAAG')