Lab Session miRNA-mRNA Networks

Daniel Gautheret, Université Paris-Sud
François Major, Université de Montréal

Retrieval of Tumor suppressors and Oncogenes from the NCG web site

Download cancer gene list from the NCG (Network of cancer Genes) web site at: http://ncg.kcl.ac.uk From the Download page, select Known cancer genes (List of known cancer genes and tumour suppressor/oncogene annotations). Read the table (tsv format) into a new data frame.

Note: we had to remove characters "#" from 3 lines, which caused read.table to fail Use function: read.table with header=T,sep="‘"

Extract oncogenes and tumor suppressors in two new vectors, based on fields *NCG6_oncogene* and *NCG6_tsg*.

Check the results.(You may use either the subset function or table[table$field==T,]).

Use function "subset" or operate on indices with: tab[tab$col1==T,]

Retrieval of miRNA/mRNA network table

Read network table *A549-control-mirbooking-with-enzymatic-efficiency.tsv* into new data frame.

Use function: read.table with header=T,sep="‘"

View the table contents in Rstudio.

Extract lines for cancer genes

Create dataframe *tsg_targets* containing data lines for all tumor suppressor genes.

Create dataframe *onc_targets* containing data lines for all oncogenes.

Display and count unique oncogenes and tumor suppressor genes in each dataframe.

Use the subset function and %in% operator. Use the unique function on the adequate column

Identifying top Oncomirs and top Tumor suppressor miRNAs.

Visualize all (target gene, miRNA) pairs in *tsg_targets*.

use table indices such as in: tab[,c("colname1","colname2")]

Note that a pair (target,miRNA) can be present several times when a miRNA has several binding sites for this target.

Aggregate all pairs (target,miRNA), while retaining for each pair the max value of column *score*. Sort result by target gene name. Check that (target,miRNA) have been correclty aggregated.

use functions: aggregate(score ~ target_name + mirna_name,onc_targets,max) sortedtab<-tab[order(tab$column),]

Now extract the most efficient (i.e. highest scoring) (target,miRNA) pairs with oncogenes and tumor suppressor as targets. This is not easy to do with base R, hence we give a solution for this:

```r
# first we order the (target,miRNA) pair table by target name and by score
oncomir<-ag_tsg[order(ag_tsg$target_name,-ag_tsg$score),]
#second we keep the first line of each target group
oncomir<-oncomir[!duplicated(oncomir$target_name),]
#finally we rank the results by overall descending score
oncomir<-oncomir[order(-oncomir$score),]

# same for tsmir:
tsmir<-ag_onc[order(ag_onc$target_name,-ag_onc$score),]
tsmir<-tsmir[!duplicated(tsmir$target_name),]
tsmir<-tsmir[order(-tsmir$score),]
```