SPINNAKER

March, 2022

Project name: SPINNAKER (SPongeINteractionNetworkmAKER)

Description: SPINNAKER (SPongeINteractionNetworkmAKER) is an R-based implementation of a methodology for identifying putative ceRNA interactions that we recently published along with its application in breast invasive carcinoma [1]. Taking as input the expression levels of RNAs and miRNAs, SPINNAKER predicts the ceRNA interaction network by implementing two modules: 1) data collection and processing, 2) ceRNA network building, deeply described in the main text.

Operating system(s): macOS High Sierra 10.13.6, Windows 10 Pro, Ubuntu 20.04.3 LTS

Programming language: R

Project page: https://github.com/sportingCode/SPINNAKER.git

Other requirements: R version 3.5.1, R 4.1.2 or higher

License: GNU AFFERO GENERAL PUBLIC LICENSE

Any restrictions to use by non-academics: license needed.

Topics documented

Requirements

All the packages required are automatically installed [1] by SPINNAKER. The user has to:

- set the own working directory in the *main.R* file (see the movie)
- set the parameters in the *config.R* file, as detailed in the following

---

`config.R` *List of the input files and parameters required to build the ceRNA network*

Description

The initial configuration requires the setting of several parameters and file names that are the input of SPINNAKER. The names of the input files must be set by the user, while the other parameters have default values that the user may change to customize the analysis.

Usage

config()

Details

The description of the parameters is specified in the following:

- 'project': name of the project folder
- 'dataset': name of the dataset folder (different datasets could exist for the same project)
- 'path': leave the default value
- 'filename_data_RNA': name of the RNA data matrix (see Input files section for details)
- 'f'ilename_data_miRNA': name of the miRNA data matrix (see Input files section for details)
- 'ceRNA1': biotype for ceRNA 1 (choose among: "ncRNA","protein-coding", "pseudo", "rRNA", "scRNA", "snoRNA", "snRNA", "tRNA") [2]
- 'ceRNA2': biotype for ceRNA 2 (choose among: "ncRNA","protein-coding", "pseudo", "rRNA", "scRNA", "snoRNA", "snRNA", "tRNA")
- 'threshold_perc_missing_values': maximum percentage of allowed missing values in the input data matrices
- 'threshold_prc_corr': threshold for the Pearson correlation coefficient (percentile) to select highly correlated ceRNA pairs
- 'threshold_prc_sensitivity': threshold for the sensitivity correlation (percentile) to select triplets
- 'searchSeedMatch': type "YES" if you want to perform the seed-match and statistical analysis in order to assign a p-value to the triplets.

[1] For macOs users (macOS 12.2.1 and R 4.1.0): if you encounter a bug when installing "gmm" package, to solve it, go to this Github release page (https://github.com/fxcoudert/gfortran-for-macOS/releases), downloaded gfortran-10.2-Catalina.dmg, and installed it. Then, you can finally load gmm package in R.

[2] SPINNAKER offers the possibility of choosing among different pool of RNAs acting as ceRNAs, as long as the total number of triplets to be tested is within the order of magnitude of $O(10^6)$, otherwise it collides with a huge computation complexity.

Example

```
#################################

project <- "TCGA"

dataset <- "brca"

path <- paste0("project/",project,"/dataset/",dataset)

#################################

# input files

filename_data_RNA <- paste0(path,"/matrix/RNA_cancer.txt")

filename_data_miRNA <- paste0(path,"/matrix/miRNA_cancer.txt")

#################################

# input parameters

ceRNA1 <- "protein-coding"

ceRNA2 <- c("ncRNA","pseudo")

threshold_perc_missing_values <- 0.10

threshold_prc_corr <- 0.99

threshold_prc_sensitivity <- 0.99

searchSeedMatch <- "YES"

#################################
```

Input files

SPINNAKER requires as input two data matrices:

- a text file providing the normalized (no raw counts) and linear (no log data) expression data for protein-coding and non-coding RNAs. The rows are the RNAs and columns are samples.
- a text file providing the normalized (no raw counts) and linear (no log data) for microRNAs mature forms. The rows are miRNAs and columns are samples (the same of RNAs).

Output files

As output, SPINNAKER creates one folder called *ceRNA network*, with three subfolders that are named *figure*, *txtFile*, and *Rdata*, detailed in the following.

1. *figure* folder contains:

   A variable number of sensitivity heatmaps (depending on the total number of RNA pairs), each one composed of a maximum number of 5000 rows (RNA pairs).

   - *heatmap_x.pdf*: it depicts the sensitivity matrix rendered as an heatmap, where rows represent the highly correlated ceRNA pairs, columns refer to all the analyzed miRNA, and sensitivity values are color-coded increasing from red to blue.

2. *txtFile* folder contains:

   - *ceRNA_interaction_newtork.txt:* a text file including the ceRNA interaction network with the triplets whose sensitivity correlation is higher than the selected threshold.
   - *ceRNA_interaction_newtork_pval.txt:* a text file including the ceRNA interaction network with the triplets whose sensitivity correlation is higher than the selected threshold, and whose p-value is statistically significant (*p-value* $\leq$ 0.05). This file is generated only if the "searchSeedMatch" parameter of config.R is set to "YES".

3. *Rdata* folder contains:

   - *data.Rdata:* output of Module 1, including the three data matrices for ceRNA type 1, ceRNA type 2, and miRNAs; and miRNA-target predictions.
   - *ceRNA_network.Rdata:* output of Module 2, including:
     i. the ceRNA interaction network with the triplets whose sensitivity correlation is higher than the selected threshold.
     ii. the ceRNA interaction network with the triplets whose sensitivity correlation is higher than the selected threshold, and whose p-value is statistically significant (*p-value* $\leq$ 0.05). This file is generated only if the "searchSeedMatch" parameter of config.R is set to "YES".
   - *parameters.Rdata*: all parameters set for the analysis.

How to cite

If you use SPINNAKER, please cite the corresponding main article.

References

[1] Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC Syst Biol* 2014; 8: 83.