getLCAi: Visualized qualitative and quantitative analysis of lung cancer aggressive index based on mRNA expression profile

Kongyongqiang

2022.04.09

Introduction

getLCAi is an R package. With the help of mRNA expression profile data, it can analyze the action direction of target gene regulation or drug treatment on lung cancer progression, so as to clearly indicate the anti-cancer or cancer promoting effect of this research factor (gene or drug).

Basic principles

Based on gene ID, the input mRNA expression profile data and background expression profile data (206 cases of expression profile data of GEO cell line) are combined and spliced. Three dimensional PCA is used to calculate the Euclidean distance, quantify the effect of processing factors, and calculate the **lung cancer progression quantitative index (lcai)**, so as to quantify and compare the function of genes or drugs. This R package plays an important role in screening lung cancer target genes and targeted drugs.

Usage and examples

1 Quick start

This R package currently contains only one function `getlcai`. If you have installed getlcai, you can use the following command to run the sample data:

```
# install
install.packages("./getLCAI_1.0.0.zip", repos = NULL, type = "win.binary")  #
Windows
install.packages("./getLCAI_1.0.0.zip", repos = NULL, type = "source")      #
Linux

library(getLCAI)
data(exp_example)
data(pheno_example)

outlist = getlcai(exp = exp_example,
                  pheno = pheno_example,
                  control = "Control",
                  experimental = "experimental",
                  type = 'Array',
                  plotPCA = TRUE)
```

This will run getlcai on the test sample data.

2 Data input

First, the `getlcai` function requires two input data:

1. mRNA expression profile data matrix
2. Sample grouping information

mRNA expression profile data matrix:

The row name is the **gene name** or the corresponding **Entrez gene ID** in NCBI, and the column name is the **sample name**. The format may be as follows:

GeneSymbol	GSM5342379	GSM5342380	GSM5342381	GSM5342382	GSM5342383	...
LOC100130938	0.994558874	0.931397019	2.125137434	1.767203536	0.941941061	...
LOC100507487	1.116883733	1.648606595	1.27224761	1.081941137	1.478542392	...
LOC145474	0.81860532	0.775973824	0.817574689	0.796898111	0.924279372	...
LOC102723721	1.768460802	0.718560512	0.693382429	0.695313642	2.204265267	...
KIAA0040	0.965309127	1.240074693	1.599558136	1.809424488	2.959996653	...
LOC101927686	1.299666438	0.989573414	1.864696408	1.562472787	2.188632799	...
...	...	...	...	...	...	...

The matrix must be **tab-delimited file** or **data Frame** (please set row name as gene and column name as sample name).

Sample grouping information:

The first column is the sample name, and the second column is the grouping information corresponding to the sample. The format has only two columns, separated by tabs. The column name is fixed and set to **sample** and **group**.

```
sample      group
GSM5342379  Control
GSM5342380  Control
GSM5342381  Control
GSM5342382  experimental
GSM5342383  experimental
GSM5342384  experimental
```

The sample grouping information must be **tab-delimited file** or **data Frame** (please set the column name to **sample** and **group**).

3 Parameter setting

Control group and experimental group

Determine which two groups to study the effect of lung cancer progression. The `control` parameter is set as the group name before treatment (control), and the `experimental` parameter is set as the group name after treatment (Experiment).

Note: do not set `control` and `experimental` to **"NSCLC"** , **"SCLC"** , **"NLE"** . These are the grouping information of background data, otherwise an error will be reported.

Source type of data

At present, the quantitative data of mRNA expression mainly comes from transcriptome **mRNA sequencing** and **chip array** technology. Based on your input data, you need to set the 'type' parameter.

1. If your expression quantitative data comes from transcriptome sequencing, please set `type` to "RNA-seq";
2. If your expression quantitative data comes from chip technology, please set `type` to **"Array"**;

Default `type = "array"` .

Whether to draw 3D PCA diagram

This R package provides a method to draw 3D PCA diagram. If you want to visualize the result data of principal component analysis, you can set `plotPCA` as **TRUE。**

4 Data output

If you run the `getlcai` command correctly, a **list** will be returned, including the following:

1. Lung cancer progression quantitative index (lcai)
2. Combined expression matrix (exp)
3. Combined grouping information (pheno)
4. The first three principal components (pcdata)

Cite

If you use this package in your published paper, please quote this paper:

XXXXXXX