

# getLCAi：基于mRNA表达谱的肺癌进展可视化定性定量分析方法

---

孔永强

2022.04.09

**getLCAi：基于mRNA表达谱的肺癌进展可视化定性定量分析方法**

[简介](#)

[基本原理](#)

[用法和示例](#)

[1 快速开始](#)

[2 数据输入](#)

[3 参数设定](#)

[4 数据输出](#)

[引用](#)

## 简介

getLCAi 是一个R包，借助mRNA表达谱数据，来分析目标基因调控或药物处理对肺癌进展的作用方向，从而明确指示该研究因素（基因或药物）抑癌或促癌作用。

## 基本原理

基于基因ID，将输入mRNA表达谱数据与背景表达谱数据（GEO细胞系表达谱数据206例）进行合并拼接，采用三维PCA计算欧式距离，量化处理因素作用的大小，计算得到**肺癌进展量化指数**（Lung Cancer Aggressive Index, LCAI），从而使基因或药物功能大小量化和可比较。本R包对筛选肺癌靶基因和靶向药物有重要作用。

## 用法和示例

### 1 快速开始

本R包目前仅包含一个函数 `getlcai`，如果您安装了 `getLCAI`，则可以使用以下命令运行示例数据：

```
● ● ●
# 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)
```

这将会在测试示例数据上运行getLCAI。

### 2 数据输入

首先，`getlcai` 函数需要两个输入数据：

1. mRNA表达矩阵
2. 样本分组信息

## mRNA表达矩阵:

行名是基因名或NCBI中对应的 Entrez Gene ID，列名是样本名，格式可能如下所示：

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	...
...	...	...	...	...	...	...

该矩阵可以为制表符分隔的文件或者data.frame (请将行名设置为基因，列名设置为样本名)。

## 样本分组信息:

第一列是样本名，第二列是样本对应的分组信息。格式只有两列，以制表符分隔，列名固定，设置为 **sample** 和 **group**。

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

该分组信息数据可以为制表符分隔的文件或者data.frame (请将列名设置为 **sample** 和 **group**)。

## 3 参数设定

### 对照组和实验组

确定研究哪两个分组间的肺癌进展效应，control 参数设置为处理前（对照）的组名，experimental 参数设置为处理后（实验）的组名。

注意：请勿将 control 和 experimental 设置为 "NSCLC", "SCLC", "NLE", 这些为背景数据的分组信息，否则将会报错。

### 数据的来源类型

目前，mRNA表达定量数据主要来源于转录组mRNA测序与芯片技术，基于您的输入数据，您需要对 type 参数进行设置。

如果您的表达定量数据来源于转录组测序，请将 type 设置为 "RNA-seq";

如果您的表达定量数据来源于芯片技术，请将 `type` 设置为 **"Array"**；

默认 `type = "Array"`。

### 是否绘制 3D PCA图

本R包提供了一种绘制 3D PCA图的方法，如果您想对主成分分析的结果数据进行可视化，可以将 `plotPCA` 设定为**TRUE**。

## 4 数据输出

正确运行 `getLCAI` 命令后，将会返回一个 `list`，包括以下内容：

1. 肺癌进展量化指数 (`lcai`)
2. 合并后的表达矩阵 (`exp`)
3. 合并后的分组信息 (`pheno`)
4. 前三个主成分 (`pcdata`)

## 引用

---

如果您在发表的论文中使用此包，请引用此论文：

XXXXXXX