Package ‘coGPS’

January 14, 2017

Type Package
Title cancer outlier Gene Profile Sets
Version 1.18.0
Date 2011-10-20
Author Yingying Wei, Michael Ochs
Maintainer Yingying Wei <ywei@jhsph.edu>
Description Gene Set Enrichment Analysis of P-value based statistics for outlier gene detection in dataset merged from multiple studies
Depends R (>= 2.13.0)
Suggests limma
Imports graphics, grDevices
License GPL-2
LazyLoad yes
biocViews Microarray, DifferentialExpression
NeedsCompilation no

R topics documented:

coGPS-package .............................................. 2
coGPS internal .............................................. 2
PatientSpecificGeneList .................................. 2
PCOPA ....................................................... 4
permCOPA ................................................... 5
PlotTopPCOPA .............................................. 6
SampleData ................................................... 8

Index 9
PatientSpecificGeneList

---

**coGPS-package**

*Cancer Outlier Gene Profile Sets*

---

**Description**

Gene Set Enrichment Analysis of P-value based statistics for outlier gene detection in dataset merged from multiple studies

**Author(s)**

Yingying Wei, Michael Ochs
Maintainer: Yingying Wei <ywei@jhsph.edu>

**References**


---

**coGPS internal**

*coGPS package internal function.*

---

**Description**

These functions are not part of the package application programming interface and are not recommended to be used by the users.

**Usage**

`plotCOPA`

---

**PatientSpecificGeneList**

*Patient Specific outlier gene list*

---

**Description**

Generate an outlier gene list for each patient restricted to the top COPA scored genes

**Usage**

`PatientSpecificGeneList(exprslist, alpha, side, type, TopGeneNum)`
Arguments

exprslist  Each element of `exprslist` is a list with the first element being `exprs` and the second element being `classlab`. Each row of `exprs` represents one gene and each column represents one sample. `classlab` is a zero-one vector indicating the status of samples. We use 0 for the baseline group, usually the normal group, and 1 for the comparison group, usually the tumor group.

alpha  Significance level for P-value.

dir  A vector specifying the definition of P-value in each of the study, which could be either `up`, `down`, or `twosided`.

type  A vector specifying whether the outlier pattern is `subtype` or `uniform`.

TopGeneNum  a number specifying the top number of outlier genes scored by PCOPA to be included in the generation of individual outlier gene list for each patient.

Value

outliergene_bypatient  a list whose length equals the number of tumor samples (patients). each element of the list is a list of length equaling to the length of `exprslist`, in other words the number of studies(or data type), showing the outlier gene for each patient in each study (or data type)

Author(s)

Yingying Wei

References


Examples

```r
#read in data
data(Exon_exprs_matched)
data(Methy_exprs_matched)
data(CNV_exprs_matched)
data(Exon_classlab_matched)
data(Methy_classlab_matched)
data(CNV_classlab_matched)
head(Exon_exprs_matched)

#exprslist[[i]]$exprs should be in matrix format
Exon_exprs<-as.matrix(Exon_exprs_matched)
Methy_exprs<-as.matrix(Methy_exprs_matched)
CNV_exprs<-as.matrix(CNV_exprs_matched)

#exprslist[[i]]$classlab should be in vector format
Exon_classlab<-unlist(Exon_classlab_matched)
Methy_classlab<-unlist(Methy_classlab_matched)
CNV_classlab<-unlist(CNV_classlab_matched)

#make an exprslist consisting 3 studies
trylist<-list()
trylist[[1]]<-list(exprs=Exon_exprs,classlab=Exon_classlab)
```
**PCOPA**

**P-value based outlier gene detection**

**Description**

Calculate P-value based statistics for outlier gene detection in dataset merged from multiple studies and give out outlier gene list for each patient.

**Usage**

```r
PCOPA(exprslist, alpha, side, type)
```

**Arguments**

- **exprslist**: Each element of `exprslist` is a list with the first element being `exprs` and the second element being `classlab`. Each row of `exprs` represents one gene and each column represents one sample. `classlab` is a zero-one vector indicating the status of samples. We use 0 for the baseline group, usually the normal group, and 1 for the comparison group, usually the tumor group.

- **alpha**: Significance level for P-value.

- **side**: A vector specifying the definition of P-value in each of the study, which could be either `up`, `down`, or `twosided`.

- **type**: A vector specifying whether the outlier pattern is `subtype` or `uniform`.

**Value**

- **PCOPAstatistics**: the P-value based outlier gene detection statistics

- **outliergene_bypatient**: a list whose length equals the number of tumor samples (patients). Each element of the list is a list of length equaling to the length of `exprslist`, in other words the number of studies (or data type), showing the outlier gene for each patient in each study (or data type)

**Author(s)**

Yingying Wei

**References**

Examples

# read in data
data(Exon_exprs_matched)
data(Methy_exprs_matched)
data(CNV_exprs_matched)
data(Exon_classlab_matched)
data(Methy_classlab_matched)
data(CNV_classlab_matched)
head(Exon_exprs_matched)

# exprslist[[i]]$exprs should be in matrix format
Exon_exprs<-as.matrix(Exon_exprs_matched)
Methy_exprs<-as.matrix(Methy_exprs_matched)
CNV_exprs<-as.matrix(CNV_exprs_matched)

# exprslist[[i]]$classlab should be in vector format
Exon_classlab<-unlist(Exon_classlab_matched)
Methy_classlab<-unlist(Methy_classlab_matched)
CNV_classlab<-unlist(CNV_classlab_matched)

# make an exprslist consisting 3 studies
trylist<-list()
trylist[[1]]<-list(exprs=Exon_exprs,classlab=Exon_classlab)
trylist[[2]]<-list(exprs=Methy_exprs,classlab=Methy_classlab)
trylist[[3]]<-list(exprs=CNV_exprs,classlab=CNV_classlab)

# calculate P-value based statistics for outlier gene detection and output the outlier gene list for each patient
a7<-permCOPA(trylist,0.05,side=c("up","down","up"),type="subtype")

permCOPA

 permCOPA Calculate PCOPA value for permutations

Description

Run permutations by randomly shuffling the sample class labels and calculate a vector of PCOPA values for each permutation.

Usage

permCOPA(exprslist, alpha=0.05, side, type, perms=100)

Arguments

eexprslist Each element of exprslist is a list with the first element being exprs and the second element being classlab. Each row of exprs represents one gene and each column represents one sample. classlab is a zero-one vector indicating the status of samples. We use 0 for the baseline group, usually the normal group, and 1 for the comparison group, usually the tumor group.

alpha Significance level for P-value.

side A vector specifying the definition of P-value in each of the study, which could be either up, down, or twosided.

type A vector specifying whether the outlier pattern is subtype or uniform.

perms Number of permutations to run.
PlotTopPCOPA

Plot expression patterns of top ranked genes.

Description

It first sorts the expression value \textit{exprslist[[i]]$exprs[j]} among the baseline samples (e.g. normal ones) and comparison group (e.g. tumor ones) separately for selected gene \textit{j}, and then plot the sorted expression values. The first argument \textit{exprslist} should be the same one as for PCOPA; the second argument \textit{PCOPAresult} should be an output of PCOPA; the third argument \textit{topcut} determines how far we would go down the top ranked list; and the last argument \textit{typelist} is a vector specifying the titles for each graph corresponds to a specific study.
Usage

PlotTopPCOPA(exprslist, PCOPAresult, topcut, typelist)

Arguments

exprslist Each element of exprslist is a list with the first element being exprs and the second element being classlab. Each row of exprs represents one gene and each column represents one sample. classlab is a zero-one vector indicating the status of samples. We use 0 for the baseline group, usually the normal group, and 1 for the comparison group, usually the tumor group.

PCOPAresult Output of PCOPA.

topcut Cutoff of top ranked gene list.

typelist A vector specifying the titles for each graph corresponds to a specific study.

Author(s)

Michael Ochs, Yingying Wei

Examples

# read in data
data(Exon_exprs_matched)
data(Methy_exprs_matched)
data(CNV_exprs_matched)
data(Exon_classlab_matched)
data(Methy_classlab_matched)
data(CNV_classlab_matched)
head(Exon_exprs_matched)

# exprslist[[i]]$exprs should be in matrix format
Exon_exprs<-as.matrix(Exon_exprs_matched)
Methy_exprs<-as.matrix(Methy_exprs_matched)
CNV_exprs<-as.matrix(CNV_exprs_matched)

# exprslist[[i]]$classlab should be in vector format
Exon_classlab<-unlist(Exon_classlab_matched)
Methy_classlab<-unlist(Methy_classlab_matched)
CNV_classlab<-unlist(CNV_classlab_matched)

# make an exprslist consisting 3 studies
trylist<-list()
trylist[[1]]<-list(exprs=Exon_exprs,classlab=Exon_classlab)
trylist[[2]]<-list(exprs=Methy_exprs,classlab=Methy_classlab)
trylist[[3]]<-list(exprs=CNV_exprs,classlab=CNV_classlab)

# calculate P-value based statistics for outlier gene detection and output the outlier gene list for each patient
a7<-PCOPA(trylist,0.05,side=c("up","down","up"),type="subtype")

# plot expression patterns of top ranked genes.
PlotTopPCOPA(trylist,a7,topcut=1,typelist=c("Exon","Methy","CNV"))
Here we present an example of coGPS analysis.

Arguments

- **Exon_exprs_matched**
  - Expression data for 44 tumors and 25 normals. Each row indicates a gene with row name showing gene name and each column indicates a sample with column name showing sample name.

- **Exon_class_matched**
  - A length 69 vector showing status of corresponding exon samples, 0 for normals and 1 for tumors.

- **Methy_exprs_matched**
  - Methylation data for 44 tumors and 25 normals.

- **Methy_class_matched**
  - A length 69 vector showing status of corresponding methylation samples, 0 for normals and 1 for tumors.

- **CNV_exprs_matched**
  - Copy number data for 44 tumors and 25 normals.

- **CNV_class_matched**
  - A length 69 vector showing status of corresponding copy number samples, 0 for normals and 1 for tumors.

- **Hs.gmtl.c1**
  - Broad Institute C1 Positional Gene Sets.

Details

In this application, the columns of each data type are matched. In other words, the first columns of Exon_exprs_matched, Methy_exprs_matched and CNV_exprs_matched correspond to the same patient. And hence the Exon_class_matched, Methy_class_matched and CNV_class_matched are identical. However, suppose in applications that we are not concerned with the outlier gene list for each patient, we can leave with the samples (columns) unmatched.
Index

*Topic Microarray, Bioinformatics, DifferentialExpression
  coGPS-package, 2

CNV_classlab_matched(SampleData), 8
CNV_exprs_matched(SampleData), 8
coGPS(coGPS-package), 2
coGPS internal, 2
coGPS-package, 2

Exon_classlab_matched(SampleData), 8
Exon_exprs_matched(SampleData), 8

Hs.gmtl.c1(SampleData), 8

Methy_classlab_matched(SampleData), 8
Methy_exprs_matched(SampleData), 8

PatientSpecificGeneList, 2
PCOPA, 4
permCOPA, 5
plotCOPA(coGPS internal), 2
PlotTopPCOPA, 6

SampleData, 8