Package ‘PRANA’

March 29, 2023

Type Package

Title Pseudo-Value Regression Approach for Network Analysis (PRANA)

Version 1.0.3

Description A novel pseudo-value regression approach for the differential co-expression network analysis in expression data, which can incorporate additional clinical variables in the model. This is a direct regression modeling for the differential network analysis, and it is therefore computationally amenable for the most users.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.1

Depends R (>= 3.5.0)

biocViews

Imports dnapath, dplyr, parallel, robustbase, stats, minet

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Language en-US

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Ahn Seungjun [cre, aut, trl] (https://orcid.org/0000-0002-4816-8924), Datta Somnath [ctb, ths]

Maintainer Ahn Seungjun <sahn1@ufl.edu>

Repository CRAN

Date/Publication 2023-03-29 08:00:09 UTC

R topics documented:

adjpval ................................................................. 2
adjpval_specific_var ............................................... 3
Description
A function to retrieve a table with adjusted p-values after running PRANA. The table includes all variables that were included in the pseudo-value regression model.

Usage
adjpval(PRANAres)

Arguments
PRANAres 
An object called after running PRANA.

Value
A table that includes adjusted p-values for all variables included in the fitted model.

Examples
data(combinedCOPDdat_RGO) # A complete data containing expression and clinical data.

# A gene expression data part of the downloaded data.
rnaseqdat = combinedCOPDdat_RGO[, 8:ncol(combinedCOPDdat_RGO)]
rnaseqdat = as.data.frame(apply(rnaseqdat, 2, as.numeric))

# A clinical data with additional covariates sorted by current smoking groups:
# The first column is ID, so do not include.
phenodat = combinedCOPDdat_RGO[order(combinedCOPDdat_RGO$currentsmoking), 2:7]

# Indices of non-current smoker (namely Group A)
index_grpA = which(combinedCOPDdat_RGO$currentsmoking == 0)
# Indices of current smoker (namely Group B)
index_grpB = which(combinedCOPDdat_RGO$currentsmoking == 1)

# Use PRANA() function to perform the pseudo-value regression analysis.
PRANAres <- PRANA(RNASeqdat = rnaseqdat, clindat = phenodat,
adjpval_specific_var

adjpval_specific_var

adjpval_specific_var

Description
A function to retrieve a vector of adjusted p-values after running PRANA.

Usage
adjpval_specific_var(adjptab, varname)

Arguments
adjptab A table that includes adjusted p-values for a specific variable.
varname Specify the name of the variable of interest.

Value
A vector of adjusted p-values for a single variable from the model.

Examples
data(combinedCOPDdat_RGO) # A complete data containing expression and clinical data.

# A gene expression data part of the downloaded data.
rnaseqdat = combinedCOPDdat_RGO[8:ncol(combinedCOPDdat_RGO)]
rnaseqdat = as.data.frame(apply(rnaseqdat, 2, as.numeric))

# A clinical data with additional covariates sorted by current smoking groups:
# The first column is ID, so do not include.
phenodat = combinedCOPDdat_RGO[order(combinedCOPDdat_RGO$currentsmoking), 2:7]

# Indices of non-current smoker (namely Group A)
index_grpA = which(combinedCOPDdat_RGO$currentsmoking == 0)
# Indices of current smoker (namely Group B)
index_grpB = which(combinedCOPDdat_RGO$currentsmoking == 1)

# Use PRANA() function to perform the pseudo-value regression analysis.
PRANAres <- PRANA(RNASeqdat = rnaseqdat, clindat = phenodat,
groupA = index_grpA, groupB = index.grpB)

# Create an object to keep the table with adjusted p-values using adjpval() function.
adjpvaltab <- adjpval(PRANAres)

# Retrieve a vector of adjusted p-values for a single variable of interest.
adjpval_specific_var(adjptab = adjpvaltab, varname = "currentsmoking")
Description

A function to retrieve a table with coefficient estimates after running PRANA. The table includes all variables that were included in the pseudo-value regression model.

Usage

```
coeff(PRANAres)
```

Arguments

- **PRANAres**: An object called after running PRANA.

Value

A table that includes coefficient estimates for all variables included in the fitted model.

Examples

```
# data(combinedCOPDdat_RGO)  # A complete data containing expression and clinical data.
# A gene expression data part of the downloaded data.
rnaseqdat = combinedCOPDdat_RGO[ , 8:ncol(combinedCOPDdat_RGO)]
rnaseqdat = as.data.frame(apply(rnaseqdat, 2, as.numeric))

# A clinical data with additional covariates sorted by current smoking groups:
# The first column is ID, so do not include.
phenodat = combinedCOPDdat_RGO[order(combinedCOPDdat_RGO$currentsmoking), 2:7]

# Indices of non-current smoker (namely Group A)
index_grpA = which(combinedCOPDdat_RGO$currentsmoking == 0)
# Indices of current smoker (namely Group B)
index_grpB = which(combinedCOPDdat_RGO$currentsmoking == 1)

# Use PRANA() function to perform the pseudo-value regression analysis.
# Then, create an object called 'res' to call results later.
PRANAres <- PRANA(RNASeqdat = rnaseqdat, clindat = phenodat,
groupA = index_grpA, groupB = index_grpB)

# Now, we want to keep the table with coefficient estimates only.
coeff(PRANAres)
```
Description

A function to retrieve a vector of coefficient estimates for a specific variable of interest after running PRANA.

Usage

coeff_specific_var(coefftab, varname)

Arguments

coefftab  A table that includes adjusted p-values for a specific variable.
varname   Specify the name of the variable of interest.

Value

A vector of coefficient estimates for a single variable from the model.

Examples

```r
# A complete data containing expression and clinical data.
data(combinedCOPDdat_RGO)

# A gene expression data part of the downloaded data.
rnaseqdat = combinedCOPDdat_RGO[, 8:ncol(combinedCOPDdat_RGO)]
rnaseqdat = as.data.frame(apply(rnaseqdat, 2, as.numeric))

# A clinical data with additional covariates sorted by current smoking groups:
# The first column is ID, so do not include.
phenodat = combinedCOPDdat_RGO[order(combinedCOPDdat_RGO$currentsmoking), 2:7]

# Indices of non-current smoker (namely Group A)
index_grpA = which(combinedCOPDdat_RGO$currentsmoking == 0)

# Indices of current smoker (namely Group B)
index_grpB = which(combinedCOPDdat_RGO$currentsmoking == 1)

# Use PRANA() function to perform the pseudo-value regression analysis.
# Then, create an object called 'res' to call results later.
PRANAres <- PRANA(RNASeqdat = rnaseqdat, clindat = phenodat,
groupA = index_grpA, groupB = index_grpB)

# Create an object to keep the table with coefficient estimates using coeff() function.
coefftab <- coeff(PRANAres)

# Lastly, we use coeff_specific_var function to retrieve
# adjusted p-values for a single variable of interest.
coeff_specific_var(coeftab = coefftab, varname = "currentsmoking")
```
**Description**

A subset of the COPDGene study data, obtained from the Gene Expression Omnibus database. This contains expression and clinical covariates data. All of the 28 genes contained in this data are identified as COPD-related by Sakornsakolpat et al. Nat Genet, 2019.

**Usage**

```r
data(combinedCOPDdat_RGO)
```

**Format**

'combinedCOPDdat_RGO'

**References**


**Examples**

```r
data(combinedCOPDdat_RGO)
combinedCOPDdat_RGO
```

---

**Description**

A function to calculate the adjusted p-value for each gene (Datta and Datta, 2005)

**Usage**

```r
EBS(pvo, alpha = 0.05, B = 500, h = 1)
```

**Arguments**

- `pvo`: An object with p-values estimated from the user-provided expression data
- `alpha`: a level of significance (default is 0.05)
- `B`: size of bootstrapping (default is 500)
- `h`: a bandwidth (default is 1)
Value

A vector of adjusted p-values for each gene.

References


Description

A pseudo-value regression approach for differential network analysis that adjusts for additional covariates (PRANA)

Usage

PRANA(RNASeqdat, clindat, groupA, groupB)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNASeqdat</td>
<td>An RNA-Seq data with subjects in rows and genes in columns.</td>
</tr>
<tr>
<td>clindat</td>
<td>A data with clinical variables to be included in the regression (e.g., binary group variable indicating current smoking status, continuous age, ...)</td>
</tr>
<tr>
<td>groupA</td>
<td>Indices of the subjects in the first category (e.g., non-current smoker) of binary group variable.</td>
</tr>
<tr>
<td>groupB</td>
<td>Indices of the subjects in the second category (e.g., current smoker) of binary group variable.</td>
</tr>
</tbody>
</table>

Value

A list containing three data frame objects that summarize the results of PRANA. This includes beta coefficients, p-values, and adjusted p-values via the empirical Bayes approach for each predictor variables that are included in the regression model.

References

Examples

data(combinedCOPDdat_RGO) # A complete data containing expression and clinical data.

# A gene expression data part of the downloaded data.
rnaseqdat = combinedCOPDdat_RGO[, 8:ncol(combinedCOPDdat_RGO)]
rnaseqdat = as.data.frame(apply(rnaseqdat, 2, as.numeric))

# A clinical data with additional covariates sorted by current smoking groups:
# The first column is ID, so do not include.
phenodat = combinedCOPDdat_RGO[order(combinedCOPDdat_RGO$currentsmoking), 2:7]

# Indices of non-current smoker (namely Group A)
index_grpA = which(combinedCOPDdat_RGO$currentsmoking == 0)
# Indices of current smoker (namely Group B)
index_grpB = which(combinedCOPDdat_RGO$currentsmoking == 1)

PRANAres <- PRANA(RNASeqdat = rnaseqdat, clindat = phenodat,
groupA = index_grpA, groupB = index_grpB)

____________________________________________________________________________________________

.sigDCGnames  .sigDCGnames

Description

A function to retrieve the name of genes that are significantly differentially connected (DC) between two biological/clinical states (aka the main binary indicator) with the presence of additional covariate information.

Usage

.sigDCGnames(adjptab, groupvar, alpha)

Arguments

adjptab A table with adjusted p-values for all variables that were included in the pseudo-value regression model.

groupvar Specify the name of binary indicator variable.

alpha A level of significance (e.g. 0.05).

Value

Names of significantly DC genes (e.g. gene IDs) from PRANA. If you need both adjusted p-values and names, please use sigDCGtab() instead.
Examples

# A complete data containing expression and clinical data.
data(combinedCOPDdat_RGO) # A gene expression data part of the downloaded data.
rnaseqdat = combinedCOPDdat_RGO[ , 8:ncol(combinedCOPDdat_RGO)]
rnaseqdat = as.data.frame(apply(rnaseqdat, 2, as.numeric))

# A clinical data with additional covariates sorted by current smoking groups:
# The first column is ID, so do not include.
phenodat = combinedCOPDdat_RGO[order(combinedCOPDdat_RGO$currentsmoking), 2:7]

# Indices of non-current smoker (namely Group A)
index grpA = which(combinedCOPDdat_RGO$currentsmoking == 0)
# Indices of current smoker (namely Group B)
index grpB = which(combinedCOPDdat_RGO$currentsmoking == 1)

# Use PRANA() function to perform the pseudo-value regression analysis.
# Then, create an object called PRANA_Results to call results.
PRANAres <- PRANA(RNASeqdat = rnaseqdat, clindat = phenodat,
groupA = index grpA, groupB = index grpB)

# Next, we want to call the table with adjusted p-values only.
adjtab <- adjpval(PRANAres)

# Please specify the name of binary group indicator in sigDCGnames(groupvar = ).
sigDCGnames <- sigDCGnames(adjtab = adjtab, groupvar = "currentsmoking", alpha = 0.05)
sigDCGnames

Description

A function to retrieve the data frame that are significantly differentially connected (DC) between
two biological/clinical states (aka the main binary indicator) with the presence of additional covariate information.

Usage

sigDCGtab(adjtab, groupvar, alpha)

Arguments

adjtab  A table with adjusted p-values and names for the variable that the user specifies in the groupvar.
groupvar  Specify the name of binary indicator variable.
alpha  A level of significance (e.g. 0.05).
Value

Adjusted p-values and names of significantly DC genes (e.g. gene IDs) from PRANA.

Examples

```
# A complete data containing expression and clinical data.
data(combinedCOPDdat_RGO)  
# A gene expression data part of the downloaded data.
rnaseqdat = combinedCOPDdat_RGO[, 8:ncol(combinedCOPDdat_RGO)]
rnaseqdat = as.data.frame(apply(rnaseqdat, 2, as.numeric))
# A clinical data with additional covariates sorted by current smoking groups:
# The first column is ID, so do not include.
phenodat = combinedCOPDdat_RGO[order(combinedCOPDdat_RGO$currentsmoking), 2:7]
index_grpA = which(combinedCOPDdat_RGO$currentsmoking == 0)
index_grpB = which(combinedCOPDdat_RGO$currentsmoking == 1)
# Use PRANA() function to perform the pseudo-value regression analysis.
# Then, create an object called PRANA_Results to call results.
PRANAres <- PRANA(RNASeqdat = rnaseqdat, clindat = phenodat, 
groupA = index_grpA, groupB = index_grpB)
# Next, we want to call the table with adjusted p-values.
adjptab <- adjpval(PRANAres)
# Please specify the name of variable in sigDCGtab(groupvar = ).
sigDCGtab <- sigDCGtab(adjptab = adjptab, groupvar = "currentsmoking", alpha = 0.05)
sigDCGtab
```

Description

A function to compute the total connectivity of each gene from the association matrix.

Usage

```
thetahats(asso.matinput)
```

Arguments

```
asso.matinput  An association matrix that is estimated from the user-provided expression data is used as an input to compute the total connectivity of each gene.
```
thetahats

Value

A vector containing total connectivity of each gene (i.e. continuous version of centrality measure of a network)
Index

* 2023.
  combinedCOPDdat_RGO, 6
* Ahn
  combinedCOPDdat_RGO, 6
* Application
  combinedCOPDdat_RGO, 6
* BMC
  combinedCOPDdat_RGO, 6
* Bioinformatics.
  combinedCOPDdat_RGO, 6
* al.
  combinedCOPDdat_RGO, 6
* data
  combinedCOPDdat_RGO, 6
* et
  combinedCOPDdat_RGO, 6
* in
  combinedCOPDdat_RGO, 6
* used
  combinedCOPDdat_RGO, 6

adjpval, 2
adjpval_specific_var, 3

coeff, 4
coeff_specific_var, 5
combinedCOPDdat_RGO, 6

EBS, 6

PRANA, 7

sigDCGnames, 8
sigDCGtab, 9

thetahats, 10