Package 'SegCorr'

October 12, 2022

Type Package

Title Detecting Correlated Genomic Regions

Version 1.2

Date 2017-12-12

Author Eleni Ioanna Delatola, Emilie Lebarbier, Tristan Mary-Huard, Francois
Radvanyi, Stephane Robin, Jennifer Wong

Maintainer Eleni Ioanna Delatola <eldelatola@yahoo.gr>

Description Performs correlation matrix segmentation and applies a test
procedure to detect highly correlated regions in gene expression.

Depends jointseg, methods

License GPL-2

RoxygenNote 5.0.0

NeedsCompilation no

Repository CRAN

Date/Publication 2017-12-17 12:49:15 UTC

R topics documented:

SegCorr-package   *Detecting Correlated Genomic Regions*

Description

Performs correlation matrix segmentation and applies a test procedure to detect highly correlated regions in gene expression. The segmentation procedure detects changes in the patterns of the gene expression correlation matrix. The test procedure asseses which regions exhibit a significantly high level of correlation. Additionally, a preprocessing procedure is provided to correct gene expression for copy number variation.

Details

Package:	SegCorr
Type:	Package
Version:	1.2
Date:	2015-01-19
License:	GPL-2

Author(s)

E. I. Delatola, E. Lebarbier, T. Mary-Huard, F. Radvanyi, S. Robin, J. Wong.

Maintainer: Eleni Ioanna Delatola <eldelatola@yahoo.gr>

References

Delatola E. I., Lebarbier E., Mary-Huard T., Radvanyi F., Robin S., Wong J.(2017). SegCorr: a statistical procedure for the detection of genomic regions of correlated expression. BMC Bioinformatics, 18:333.

See Also

Fpsn

Examples

```
#data.sets = c('SNP','EXP_raw')
## Each gene corresponds to one SNP probe ##
#Position_EXP = matrix(1:1000,nrow=500,byrow=TRUE)
#Position_SNP = seq(2,1000,by=2)
#data(list=data.sets)
#CHR = rep(1,dim(EXP_raw)[1])
#SNP.CHR = rep(1,dim(SNP)[1])
```

```
#results = SegCorr(CHR = CHR, EXP = EXP_raw, CNV = TRUE, SNPSMOOTH=TRUE,
#Position.EXP = Position_EXP, SNP.CHR = SNP.CHR, SNP=SNP , Position.SNP = Position_SNP)

###############drawing the heatmap for one region #########################
#tau = results$Region.List[1,2]: results$Region.List[1,3]
#EXP.CNV =  results$EXP.corrected
#heatmap(EXP.CNV[tau,])
```

CNV_correction	*Corrects Gene Expression for CNV*

Description

Correcting gene expression signal for CNV.

Usage

```
CNV_correction(s.Position.EXP, e.Position.EXP, Position.SNP, mu.SNP, EXP)
```

Arguments

`s.Position.EXP`	vector with gene start position
`e.Position.EXP`	vector with gene end position
`Position.SNP`	vector with SNP/CGH positions
`mu.SNP`	Smoothed genomic signal matrix not containing NA values. Rows correspond to probes, while columns to patients. The ordering of the patients must be the same as in the EXP matrix.
`EXP`	Gene expression matrix must not contain NA's and genes with same expression value (i.e. null gene). Rows correspond to probes, while columns to patients. Again, ordering of patients must be the same between EXP and mu.SNP matrices.

Details

Overlapping genes may correspond to the same SNP/CGH probes.

Value

CNV corrected signal matrix.

Author(s)

E. I. Delatola, E. Lebarbier, T. Mary-Huard, F. Radvanyi, S. Robin, J. Wong.

References

Delatola E. I., Lebarbier E., Mary-Huard T., Radvanyi F., Robin S., Wong J.(2017). SegCorr: a statistical procedure for the detection of genomic regions of correlated expression. BMC Bioinformatics, 18:333.

See Also

segmented_signal

Examples

```
#data.sets = c('SNP','EXP_raw')
## Each gene corresponds to one SNP probe ##
#Position_EXP = matrix(1:1000,nrow=500,byrow=TRUE)
#Position_SNP = seq(2,1000,by=2)
#data(list=data.sets)
#mu.SNP = segmented_signal(SNP ,100) ## smoothed SNP signal
#EXP.CNV = CNV_correction(Position_EXP[,1], Position_EXP[,2], Position_SNP,
#mu.SNP, EXP_raw)## corrected signal
```

EXP_raw   *Simulated Gene Expression*

Description

Gene expression profiles have been generated for 30 patients and 500 genes. Background correlation is set to 0.08 and the correlation for H1 regions to 0.5. The location of the H1 regions is as suggested in the work of Lai et al.(2005), i.e region 1 [101, 105], region 2 [201, 210], region 3 [301, 320] and region 4 [401, 440].

Usage

```
data("EXP_raw")
```

Format

Data frame containing the gene expression signal for 500 genes (rows) on 30 patients (columns).

References

Lai, W. R., Johnson, M. D., Kucherlapati, R., & Park, P. J. (2005). Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. Bioinformatics, 21(19), 3763-3770.

Examples

```
#data(EXP_raw)
#G = cor(t(EXP_raw))## calculating the gene x gene correlation matrix
#image(G)## plotting the correlation matrix
```

SegCorr	*Performs CNV Correction and Correlation Matrix Segmentation*

Description

Gene expression is corrected for CNV events must not contain NA's and genes with same expression value (i.e. null gene expression). Segmentation is used to detect changes in the correlation pattern. Regions with high correlation are identified using an exact test.

Usage

```
SegCorr(CHR, EXP, genes,S, CNV, SNPSMOOTH, Position.EXP, SNP.CHR, SNP, Position.SNP, Kmax)
```

Arguments

CHR	Chromosome allocation vector for the genes.
EXP	Gene expression matrix (raw/corrected for CNV). Columns correspond to patients and rows to genes. The expression matrix must not contain either NA's or genes with same expression value (i.e. null gene expression)
genes	Gene ID(name) vector.
S	Threshold for model selection. Default S=0.7.
CNV	Logical variable indicating whether to perform CNV correction. When CNV=T, the correction is performed. Default value CNV=F.
SNPSMOOTH	(Optional Argument when CNV=T) Logical variable indicating whether to perform SNPSMOOTH. When SNPSMOOTH=T, the smoothing is performed. Default value SNPSMOOTH=F.
Position.EXP	(Optional Argument when CNV=T) Expression position matrix. First column is the start position and the second is the end position.
SNP.CHR	(Optional Argument when CNV=T) Chromosome allocation vector for genomic probes.
SNP	(Optional Argument when CNV=T) SNP profile matrix not containing NA's. Columns correspond to patients and rows to probes.
Position.SNP	(Optional Argument when CNV=T) vector with SNP positions
Kmax	(Optional Argument when CNV=T and SNPSMOOTH=T) Maximum number of segments. (mean profile segmentation)

Details

Overlapping genes may correspond to the same genomic probes.

Value

Results	Matrix containing information about the genomic regions. Each region corresponds to a row of the matrix, the one with the smallest p-value is on the top of the list.
Results$CHR	Chromosome

Results$Start/End

the region boundaries with repsect to the physical location of the gene in the chromosome

Results$Rho	ρ correlation
Results$length	number of genes in the region

Results$first/last gene

name of the first/last gene in the region

Results$p-value

p-value as obtained from the test

Results$genes   names of the genes belonging to the region

Results$p-valueadj

p-value of the region corrected for multiple testing

Chromosome.Inf	Matrix containing the estimated background correlation (rho0.hat) per chromsome, the number of segments and the log-loglikehood.
EXP.corrected	If the CNV option is chosen, the corrected signal is given.

Author(s)

E. I. Delatola, E. Lebarbier, T. Mary-Huard, F. Radvanyi, S. Robin, J. Wong.

References

Delatola E. I., Lebarbier E., Mary-Huard T., Radvanyi F., Robin S., Wong J.(2017). SegCorr: a statistical procedure for the detection of genomic regions of correlated expression. BMC Bioinformatics, 18:333.

See Also

CNV_correction,segmentation

Examples

```
#data('EXP_raw')
#CHR = rep(1,dim(EXP_raw)[1])

#results = SegCorr(CHR = CHR, EXP = EXP_raw, CNV = FALSE,S=0.7)

###############drawing the heatmap for one region ##########################
#tau = results$Region.List[1,2]: results$Region.List[1,3]
#heatmap(as.matrix(EXP_raw[tau,]))
```

---

segmentation   *Correlation Matrix Segmentation*

Description

For a given chromosome, gene correlation matrix segmentation is performed. Regions with high correlation are identified using an exact test. The expression matrix must not contain NA's and genes with same expression value (i.e. null gene expression).

Usage

```
segmentation(CHR, EXP, genes, S)
```

Arguments

CHR	chromosome name
EXP	Gene expression matrix (raw/corrected for CNV). Columns correspond to patients and rows to genes. The expression matrix must not contain either NA's or genes with same expression value (i.e. null gene expression).
genes	Gene ID(name) vector.
S	Threshold for model selection. Default S=0.7.

Value

Results	Matrix containing information about the genomic regions. Each region corresponds to a row of the matrix, the one with the smallest p-value is on the top of the list.
Results$CHR	Chromosome
Results$Start/End	
	region boundaries with respect to the physical location of the gene in the chromosome
Results$Rho	ρ correlation
Results$length	number of genes in the region
Results$first/last gene	
	name of the first/last gene in the region
Results$p-value	
	p-value as obtained from the test
Results$genes	names of genes belonging to the region
rho0	estimate of the background correlation
likelihood	log-likelihood
K	number of segments

Author(s)

E. I. Delatola, E. Lebarbier, T. Mary-Huard, F. Radvanyi, S. Robin, J. Wong.

References

Delatola E. I., Lebarbier E., Mary-Huard T., Radvanyi F., Robin S., Wong J.(2017). SegCorr: a statistical procedure for the detection of genomic regions of correlated expression. BMC Bioinformatics, 18:333.

Examples

```
#data(EXP_raw)
#G = cor(t(EXP_raw))## calculating the gene x gene correlation matrix
#image(G)## plotting the correlation matrix
#results = segmentation(EXP = EXP_raw)
```

---

segmented_signal   *Mean Segmentation*

Description

Mean segmentation on the genomic signal is performed using the Fpsn function of the jointseg package.

Usage

```
segmented_signal(SNP.Chr, Kmax)
```

Arguments

SNP.Chr	SNP/CGH profile matrix for a given chromosome (NA's not allowed). Columns correspond to patients and rows to probes.
Kmax	Maximum number of segments.

Value

Smoothed genomic signal matrix. Rows correspond to probes and columns to patients.

Author(s)

E. I. Delatola, E. Lebarbier, T. Mary-Huard, F. Radvanyi, S. Robin, J. Wong.

References

Morgane Pierre-Jean, Guillem Rigaill and Pierre Neuvial. Performance evaluation of DNA copy number segmentation methods. Briefings in Bioinformatics (2015) 16 (4): 600-615.

See Also

[CNV_correction](), [Fpsn]()

Examples

```
#data(SNP)
#mu.SNP = segmented_signal(SNP ,100)
```

SNP   *Simulated SNP signal*

Description

SNP profiles for 30 patients and 500 probes have been simulated. Each gene corresponds to one SNP value.

Usage

```
data("SNP")
```

Format

Data frame with 500 probes (rows) on 30 patients (patients).

Examples

```
#data(SNP)
```

Index