

Package ‘TSEA’

August 20, 2018

Type Package

Title Tissue-Specific Enrichment Analysis

Version 1.0

Date 2018-08-20

Author Guangsheng Pei

Maintainer Guangsheng Pei <guangsheng.pei@uth.tmc.edu>

Imports pheatmap (>= 1.0.10), RColorBrewer (>= 1.1)

Description Tissue-specific enrichment analysis to assess lists of candidate genes or RNA-Seq expression profiles.

License GPL (>= 2)

NeedsCompilation no

Depends R (>= 2.10)

R topics documented:

TSEA-package	2
correction_factor	5
ENCODE_z_score	6
GTEEx_t_score	6
GWAS_gene	7
GWAS_gene_multiple	8
query_ENCODE	9
query_GTEEx	11
tsea.analysis	11
tsea.analysis.multiple	12
tsea.expression.decode	13
tsea.expression.normalization	14
tsea.plot	15
tsea.summary	16
Index	18

TSEA-package	<i>Tissue-Specific Enrichment Analysis Tissue-Specific Enrichment Analysis</i>
--------------	--

Description

Tissue-specific enrichment analysis to assess lists of candidate genes and tissue-specific expression decode analysis for RNA-seq data to decode RNA expression matrices tissue heterogeneity.

Details

Since disease and physiological condition are often associated with a specific tissue, understanding the tissue-specific genes (TSG) expression patterns will substantially reduce false discoveries in biomedical research. However, due to cell complexity in human system, heterogeneous tissues are frequently collected. Making it difficult to distinguish gene expression variability and mislead result interpretation. Here, we present TSEA, an R package that conducts Tissue-Specific Enrichment Analysis (TSEA) using two built-in reference panels: the Genotype-Tissue Expression (GTEx) data and the ENCyclopedia Of DNA Elements (ENCODE) data. We implemented two major functions in TSEA to assess lists of candidate genes or expression matrices.

The DESCRIPTION file:

```
Package:      TSEA
Type:         Package
Title:        Tissue-Specific Enrichment Analysis
Version:      1.0
Date:         2018-08-20
Author:       Guangsheng Pei
Maintainer:   Guangsheng Pei <guangsheng.pei@uth.tmc.edu>
Imports:      pheatmap (>= 1.0.10), RColorBrewer (>= 1.1)
Description:  Tissue-specific enrichment analysis to assess lists of candidate genes or RNA-Seq expression profiles.
License:      GPL (>= 2)
```

Index of help topics:

ENCODE_z_score	ENCODE z-score to define tissue-specific genes
GTEx_t_score	GTEx t-score to define tissue-specific genes
GWAS_gene	Gene symbol query data for single sample
GWAS_gene_multiple	Gene symbol query data for multiple samples
TSEA-package	Tissue-Specific Enrichment Analysis
	Tissue-Specific Enrichment Analysis
correction_factor	Gene average expression level and standard deviation in GTEx data
query_ENCODE	ENCODE raw query data
query_GTEx	GTEx raw query data
tsea.analysis	Tissue-specific enrichment analysis for query gene list
tsea.analysis.multiple	Tissue-specific enrichment analysis for multi query gene lists

tsea.expression.decode	Tissue-specific enrichment analysis for RNA-Seq expression profiles
tsea.expression.normalization	RNA-Seq expression profiles normalization
tsea.plot	Tissue-specific enrichment analysis result heatmap plot
tsea.summary	Tissue-specific enrichment analysis result summary

Author(s)

Guangsheng Pei

Maintainer: Guangsheng Pei

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

See Also

<https://github.com/bsml320/TSEA/>

Examples

```
data(GTEx_t_score)
data(ENCODE_z_score)
library(pheatmap)

#=====
#Example 1: Lists of candidate genes for single sample.
data(GWAS_gene)
query_gene_list = GWAS_gene

#TSEA in GTEx panel
tsea_t = tsea.analysis(query_gene_list, GTEx_t_score, 0.05,
p.adjust.method = "bonferroni")
#write.csv(tsea_t,"1.1.GWAS_TSEA_in_GTEx_panel.csv")
#pdf ("1.2.GWAS_TSEA_in_GTEx_panel.pdf", 5, 9, onefile = FALSE)
#tsea.plot(tsea_t, 0.05)
#dev.off()
tsea_t_summary = tsea.summary(tsea_t)
#write.csv(tsea_t_summary,"1.3.GWAS_summary_in_GTEx_panel.csv")

#TSEA in ENCODE panel
tsea_z = tsea.analysis(query_gene_list, ENCODE_z_score, 0.05,
p.adjust.method = "bonferroni")
#write.csv(tsea_z,"1.4.GWAS_TSEA_in_ENCODE_panel.csv")
#pdf ("1.5.GWAS_TSEA_in_ENCODE_panel.pdf", 5, 9, onefile = FALSE)
#tsea.plot(tsea_z, 0.05)
#dev.off()
tsea_z_summary = tsea.summary(tsea_z)
#write.csv(tsea_z_summary,"1.6.GWAS_summary_in_ENCODE_panel.csv")

#=====
```

```

#Example 2: Lists of candidate genes for multiple samples.
data(GWAS_gene_multiple)
query_gene_list = GWAS_gene_multiple

#TSEA in GTEx panel
tsea_t_multi = tsea.analysis.multiple(query_gene_list,
GTEx_t_score, 0.05, p.adjust.method = "BH")
#write.csv(tsea_t_multi,"2.1.GWAS_multi_TSEA_in_GTEX_panel.csv")
#pdf ("2.2.GWAS_multi_TSEA_in_GTEX_panel.pdf", 6, 6, onefile = FALSE)
#tsea.plot(tsea_t_multi, 0.05)
#dev.off()
tsea_t_multi_summary = tsea.summary(tsea_t_multi)
#write.csv(tsea_t_multi_summary,"2.3.GWAS_multi_summary_in_GTEX_panel.csv")

#TSEA in ENCODE panel
tsea_z_multi = tsea.analysis.multiple(query_gene_list,
ENCODE_z_score, 0.05, p.adjust.method = "BH")
#write.csv(tsea_z_multi,"2.4.GWAS_multi_TSEA_in_ENCODE_panel.csv")
#pdf ("2.5.GWAS_multi_TSEA_in_ENCODE_panel.pdf", 7, 7, onefile = FALSE)
#tsea.plot(tsea_z_multi, 0.2)
#dev.off()
tsea_z_multi_summary = tsea.summary(tsea_z_multi)
#write.csv(tsea_z_multi_summary,"2.6.GWAS_multi_summary_in_ENCODE_panel.csv")

#=====
#Example 3: RNA expression profiles TSEA in ENCODE panel.
data(query_GTEx)
query_matrix = query_GTEx

data(correction_factor)
#RNA expression profiles z-score normalization
query_mat_zscore_nor = tsea.expression.normalization(query_matrix,
correction_factor, normalization = "z-score")

#RNA expression profiles TSEA in ENCODE panel
tseaed_in_ENCODE = tsea.expression.decode(query_mat_zscore_nor,
ENCODE_z_score, 0.05, p.adjust.method = "BH")
#write.csv(tseaed_in_ENCODE,"3.1.RNAseq_TSEA_in_ENCODE_panel.csv")
#pdf ("3.2.RNAseq_TSEA_in_ENCODE_panel.pdf", 10, 9, onefile = FALSE)
#tsea.plot(tseaed_in_ENCODE, 0.05)
#dev.off()
tseaed_in_ENCODE_summary = tsea.summary(tseaed_in_ENCODE)
#write.csv(tseaed_in_ENCODE_summary,"3.3.RNAseq_summary_in_ENCODE_panel.csv")

#=====
#Example 4: RNA expression profiles TSEA in GTEx panel.
data(query_ENCODE)
query_matrix = query_ENCODE

#RNA expression profiles abundance normalization
query_mat_abundance_nor = tsea.expression.normalization(query_matrix,
correction_factor, normalization = "abundance")

#RNA expression profiles TSEA in GTEx panel
tseaed_in_GTEx = tsea.expression.decode(query_mat_abundance_nor,
GTEx_t_score, 0.05, p.adjust.method = "BH")
#write.csv(tseaed_in_GTEx,"4.1.RNAseq_TSEA_in_GTEX_panel.csv")

```

```
#pdf ("4.2.RNAseq_TSEA_in_GTEX_panel.pdf", 10, 9, onefile = FALSE)
#tsea.plot(tseaed_in_GTEX, 0.05)
#dev.off()
tseaed_in_GTEX_summary = tsea.summary(tseaed_in_GTEX)
#write.csv(tseaed_in_GTEX_summary, "4.3.RNAseq_summary_in_GTEX_panel.csv")
```

correction_factor	<i>Gene average expression level and standard deviation in GTEX data</i>
-------------------	--

Description

Gene average expression level and standard deviation in GTEX data

Usage

```
data("correction_factor")
```

Format

A data frame with 14725 observations on the following 2 variables.

```
avg.all  a factor with levels 0 0.000112140530167417 0.000114315778302239 0.000123995511284
sd.all   a factor with levels 0 0.000101374794387476 0.000114808260468376 0.0001316311420376
```

Details

nothing

Source

nothing

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

Examples

```
data(correction_factor)
```

ENCODE_z_score	<i>ENCODE z-score to define tissue-specific genes</i>
----------------	---

Description

ENCODE z-score matrix to define tissue-specific genes

Usage

```
data("ENCODE_z_score")
```

Format

The format is: num [1:14031, 1:44] -0.6738 -0.0779 -0.0929 3.0283 -0.6371 ... - attr(*, "dimnames")=List of 2 ..\$: chr [1:14031] "C1orf112" "FGR" "CFH" "FUCA2"\$: chr [1:44] "Adrenal Gland" "Body of Pancreas" "Breast Epithelium" "Camera-type Eye" ...

Details

nothing

Source

nothing

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

Examples

```
data(ENCODE_z_score)
```

GTEx_t_score	<i>GTEx t-score to define tissue-specific genes</i>
--------------	---

Description

GTEx t-score matrix to define tissue-specific genes

Usage

```
data("GTEx_t_score")
```

Format

The format is: num [1:14725, 1:47] -0.524 -9.921 -6.812 -7.785 -7.113 ... - attr(*, "dimnames")=List of 2 ..\$: chr [1:14725] "OR4F5" "SAMD11" "KLHL17" "PLEKHN1"\$: chr [1:47] "Adipose - Subcutaneous" "Adipose - Visceral (Omentum)" "Adrenal Gland" "Artery - Aorta" ...

Details

nothing

Source

nothing

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

Examples

```
data(GTEx_t_score)
```

GWAS_gene	<i>Gene symbol query data for single sample</i>
-----------	---

Description

An example of input gene symbol query data for single sample tissue-specific enrichment analysis

Usage

```
data("GWAS_gene")
```

Format

The format is: chr [1:21729] "A1BG" "A1BG-AS1" "A1CF" "A2M" "A2M-AS1" "A2ML1" "A2MP1" "A3GALT2" "A4GALT" "A4GNT" "AA06" "AAAS" "AACS" "AACSP1" "AADAC" "AADACL2" ...

Details

nothing

Source

nothing

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

Examples

```
data(GWAS_gene)
```

GWAS_gene_multiple *Gene symbol query data for multiple samples*

Description

An example of input gene symbol query data for multiple samples tissue-specific enrichment analysis

Usage

```
data("GWAS_gene_multiple")
```

Format

A data frame with 21729 observations on the following 26 variables.

ALZ a numeric vector
ADHD a numeric vector
ASD a numeric vector
BD a numeric vector
MDD a numeric vector
SCZ a numeric vector
BMI a numeric vector
‘FN-BMD’ a numeric vector
‘LS-BMD’ a numeric vector
EDU a numeric vector
HEIGHT a numeric vector
WHR a numeric vector
CD a numeric vector
IBD a numeric vector
RA a numeric vector
UC a numeric vector
AAM a numeric vector
CAD a numeric vector
FG a numeric vector
FI a numeric vector
HDL a numeric vector
LDL a numeric vector
TC a numeric vector
TG a numeric vector
T1D a numeric vector
T2D a numeric vector

Details

nothing

Source

nothing

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

Examples

```
data(GWAS_gene_multiple)
```

query_ENCODE	<i>ENCODE raw query data</i>
--------------	------------------------------

Description

An example of RNA-Seq query data from ENCODE data for tissue-specific enrichment analysis

Usage

```
data("query_ENCODE")
```

Format

A data frame with 18661 observations on the following 44 variables.

‘Adrenal Gland’ a numeric vector
 ‘Body of Pancreas’ a numeric vector
 ‘Breast Epithelium’ a numeric vector
 ‘Camera-type Eye’ a numeric vector
 Cerebellum a numeric vector
 Diencephalon a numeric vector
 ‘Esophagus Muscularis Mucosa’ a numeric vector
 ‘Esophagus Squamous Epithelium’ a numeric vector
 ‘Frontal Cortex’ a numeric vector
 ‘Gastrocnemius Medialis’ a numeric vector
 ‘Gastroesophageal Sphincter’ a numeric vector
 Heart a numeric vector
 ‘Heart Left Ventricle’ a numeric vector
 Liver a numeric vector
 ‘Lower Leg Skin’ a numeric vector
 Lung a numeric vector

Metanephros a numeric vector
'Occipital Lobe' a numeric vector
'Omental Fat Pad' a numeric vector
Ovary a numeric vector
'Parietal Lobe' a numeric vector
'Peyer's Patch' a numeric vector
'Prostate Gland' a numeric vector
'Right Atrium Auricular Region' a numeric vector
'Right Lobe of Liver' a numeric vector
'Sigmoid Colon' a numeric vector
'Skeletal Muscle Tissue' a numeric vector
'Skin of Body' a numeric vector
'Spinal Cord' a numeric vector
Spleen a numeric vector
Stomach a numeric vector
'Subcutaneous Adipose Tissue' a numeric vector
'Suprapubic Skin' a numeric vector
'Temporal Lobe' a numeric vector
Testis a numeric vector
'Thoracic Aorta' a numeric vector
'Thyroid Gland' a numeric vector
'Tibial Nerve' a numeric vector
Tongue a numeric vector
'Transverse Colon' a numeric vector
'Umbilical Cord' a numeric vector
'Upper Lobe of Left Lung' a numeric vector
'Urinary Bladder' a numeric vector
Uterus a numeric vector

Details

nothing

Source

nothing

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

Examples

```
data(query_ENCODE)
```

query_GTEx

GTEx raw query data

Description

An example of RNA-Seq query data from GTEx data for tissue-specific enrichment analysis

Usage

```
data("query_GTEx")
```

Format

The format is: num [1:18067, 1:49] 0.0317 0.4451 21.9084 4.1406 0.4531 ... - attr(*, "dim-names")=List of 2 ..\$: chr [1:18067] "OR4F5" "SAMD11" "NOC2L" "KLHL17"\$: chr [1:49] "Adipose - Subcutaneous" "Adipose - Visceral (Omentum)" "Adrenal Gland" "Artery - Aorta" ...

Details

nothing

Source

nothing

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

Examples

```
data(query_GTEx)
```

tsea.analysis

Tissue-specific enrichment analysis for query gene list

Description

Tissue-specific enrichment analysis by Fisher's Exact Test for given gene list.

Usage

```
tsea.analysis(query_gene_list, score = GTEx_t_score, ratio = 0.05, p.adjust.meth
```

Arguments

`query_gene_list` a gene symbol list object.

`score` a gene tissue-specific score matrix, c("GTEx_t_score" or "ENCODE_z_score"), can be loaded by data(GTEx) or data(ENCODE), the default value is "GTEx_t_score".

`ratio` the threshold to define tissue-specific genes (with top t-score or z-score), the default value is 0.05.

`p.adjust.method` p.adjust.method, c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none")

Details

Tissue-specific enrichment analysis by Fisher's Exact Test for given gene list.

Value

nothing

Note

nothing

Author(s)

Guangsheng Pei

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

See Also

<https://github.com/bsml320/TSEA/>

tsea.analysis.multiple

Tissue-specific enrichment analysis for multi query gene lists

Description

Tissue-specific enrichment analysis by Fisher's Exact Test for multiple gene list.

Usage

```
tsea.analysis.multiple(query_gene_list, score = GTEx_t_score, ratio = 0.05,
p.adjust.method = "BH")
```

Arguments

<code>query_gene_list</code>	a 0~1 gene~sample table object, row should be gene symbol, column should be sample name. In the table, gene labeled with 1 indicated it is target gene for a given sample, while 0 indicated it is not target in a given sample.
<code>score</code>	a gene tissue-specific score matrix, c("GTEx_t_score" or "ENCODE_z_score"), can be loaded by data(GTEx) or data(ENCODE), the default value is "GTEx_t_score".
<code>ratio</code>	the threshold to define tissue-specific genes (with top t-score or z-score), the default value is 0.05.
<code>p.adjust.method</code>	p.adjust.method, c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none")

Details

Tissue-specific enrichment analysis by Fisher's Exact Test for multiple gene list.

Value

nothing

Note

nothing

Author(s)

Guangsheng Pei

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

See Also

<https://github.com/bsml320/TSEA/>

tsea.expression.decode

Tissue-specific enrichment analysis for RNA-Seq expression profiles

Description

Tissue-specific enrichment analysis to decode whether a given RNA-seq sample (RPKM) with potential confounding effects based on expression profiles.

Usage

```
tsea.expression.decode(query_mat_normalized_score, score = GTEx_t_score, ratio =
p.adjust.method = "BH")
```

Arguments

<code>query_mat_normalized_score</code>	a normalized RNA-seq RPKM object, which produced by "tsea.expression.normalization".
<code>score</code>	a gene tissue-specific score matrix, c("GTEx_t_score" or "ENCODE_z_score"), can be loaded by data(GTEx) or data(ENCODE), the default value is "GTEx_t_score".
<code>ratio</code>	the threshold to define tissue-specific genes (with top t-score or z-score), the default value is 0.05.
<code>p.adjust.method</code>	p.adjust.method, c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none")

Details

Tissue-specific enrichment analysis for RNA-Seq expression profiles.

Value

nothing

Note

nothing

Author(s)

Guangsheng Pei

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

See Also

<https://github.com/bsml320/TSEA/>

tsea.expression.normalization

RNA-Seq expression profiles normalization

Description

To avoid the data bias and adapt better data heterogeneity, before tsea.expression.decode() analysis, the raw discrete RPKM value have to normalized to continuous variable meet the normal distribution before t-test.

Usage

```
tsea.expression.normalization(query_mat, correction_factor = correction_factor,
normalization = "abundance")
```

Arguments

`query_mat` a RNA-seq RPKM object, row name should be gene symbol, and column name should be sample name.

`correction_factor` `correction_factor`, a gene table object contain genes average expression level and standard variance in GTEx database, can be loaded by `data(correction_factor)`.

`normalization` normalization methods, `c("z-score", "abundance")`

Details

As RNA-Seq samples are often heterogeneous, before in-depth analysis, it is necessary to decode tissue heterogeneity to avoid samples with confounding effects. However, the raw discrete RPKM value have to normalized to continuous variable meet the normal distribution before t-test.

Value

nothing

Note

nothing

Author(s)

Guangsheng Pei

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

See Also

<https://github.com/bsml320/TSEA/>

tsea.plot

Tissue-specific enrichment analysis result heatmap plot

Description

Heat map plot for tissue-specific enrichment analysis result.

Usage

```
tsea.plot(tsea_result, threshold = 0.05)
```

Arguments

- `tsea_result` the result of tissue-specific enrichment analysis, which produced by "tsea.analysis", "tsea.analysis.multiple" or "tsea.expression.decode".
- `threshold` the p-value threshold to define if the gene list or RNA-seq data enriched in a given tissue, p-value greater than threshold will not be labeled in the plot. The default value is 0.05.

Details

Heat map plot for tissue-specific enrichment analysis result

Value

nothing

Note

nothing

Author(s)

Guangsheng Pei

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

See Also

<https://github.com/bsml320/TSEA/>

`tsea.summary`

Tissue-specific enrichment analysis result summary

Description

Tissue-specific enrichment analysis result summary (list the top 3 most enriched tissues) from the given gene list or RNA-seq expression profiles.

Usage

```
tsea.summary(tsea_result)
```

Arguments

- `tsea_result` the result of tissue-specific enrichment analysis, which produced by "tsea.analysis", "tsea.analysis.multiple" or "tsea.expression.decode".

Details

Tissue-specific enrichment analysis result summary

Value

nothing

Note

nothing

Author(s)

Guangsheng Pei

References

Pei G., Dai Y., Zhao Z. Jia P. (2018) Tissue-Specific Enrichment Analysis (TSEA) to decode tissue heterogeneity. Bioinformatics, In submission.

See Also

<https://github.com/bsml320/TSEA/>

Index

***Topic TSEA** tsea.summary, 16
TSEA-package, 2

***Topic \textasciitildekw1**
tsea.analysis, 11
tsea.analysis.multiple, 12
tsea.expression.decode, 13
tsea.expression.normalization, 14
tsea.plot, 15
tsea.summary, 16

***Topic \textasciitildekw2**
tsea.analysis, 11
tsea.analysis.multiple, 12
tsea.expression.decode, 13
tsea.expression.normalization, 14
tsea.plot, 15
tsea.summary, 16

***Topic datasets**
correction_factor, 5
ENCODE_z_score, 6
GTEx_t_score, 6
GWAS_gene, 7
GWAS_gene_multiple, 8
query_ENCODE, 9
query_GTEx, 11

correction_factor, 5

ENCODE_z_score, 6

GTEx_t_score, 6
GWAS_gene, 7
GWAS_gene_multiple, 8

query_ENCODE, 9
query_GTEx, 11

TSEA (*TSEA-package*), 2
TSEA-package, 2
tsea.analysis, 11
tsea.analysis.multiple, 12
tsea.expression.decode, 13
tsea.expression.normalization, 14
tsea.plot, 15