Package 'TSEA'

August 22, 2018

Type Package

Title Tissue-Specific Enrichment Analysis

Version 1.0

Date 2018-08-21

Author Guangsheng Pei

Maintainer Guangsheng Pei <peiguangsheng@gmail.com>

Imports pheatmap, RColorBrewer

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

License GPL (>= 2)

NeedsCompilation no

R topics documented:

TSEA-package	*Tissue-Specific Enrichment Analysis Tissue-Specific Enrichment Analysis*

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.

1

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:	2018-08-21
Author:	Guangsheng Pei
Maintainer:	Guangsheng Pei <peiguangsheng@gmail.com>
Imports:	pheatmap, RColorBrewer
Description:	Tissue-specific enrichment analysis to assess lists of candidate genes or RNA-Seq expression profiles.
License:	GPL (>= 2)

Index of help topics:

```
TSEA-package            Tissue-Specific Enrichment Analysis
                        Tissue-Specific Enrichment Analysis
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 specificity. Bioinformatics, In submission.

See Also

https://github.com/bsml320/TSEA/blob/master/README.md

Examples

```
#Download the built-in data from https://github.com/bsml320/TSEA/tree/master/data
#load("data/GTEx_t_score.rda")
#load("data/ENCODE_z_score.rda")
#library(pheatmap)

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

#TSEA in GTEx panel
#tsea_t = tsea.analysis(query_gene_list, GTEx_t_score, ratio = 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, threshold = 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, ratio = 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, threshold = 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.
#load("data/GWAS_gene_multiple.rda")
#query_gene_list = GWAS_gene_multiple

#TSEA in GTEx panel
#tsea_t_multi = tsea.analysis.multiple(query_gene_list,
#GTEx_t_score, ratio = 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, threshold = 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, ratio = 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, threshold = 0.05)
```

```
#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.
#load("data/query_GTEx.rda")
#query_matrix = query_GTEx

#load("data/correction_factor.rda")
#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, ratio = 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, threshold = 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.
#load("data/query_ENCODE.rda")
#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, ratio = 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, threshold = 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")
```

| 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, ratio = 0.05,
p.adjust.method = "BH")
```

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 recommended "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 specificity. Bioinformatics, In submission.

See Also

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

Examples

```
#load("data/GWAS_gene.rda")
#query_gene_list = GWAS_gene
#tsea_t = tsea.analysis(query_gene_list, GTEx_t_score, 0.05,
#p.adjust.method = "bonferroni")
```

---

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, ratio = 0.05,
p.adjust.method = "BH")
```

Arguments

query_gene_list

> 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.

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 recommended "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 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 specificity. Bioinformatics, In submission.

See Also

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

Examples

```
#load("data/GWAS_gene_multiple.rda")
#query_gene_list = GWAS_gene_multiple
#tsea_t_multi = tsea.analysis.multiple(query_gene_list,
#GTEx_t_score, 0.05, p.adjust.method = "BH")
```

```
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,
ratio = 0.05, p.adjust.method = "BH")
```

Arguments

query_mat_normalized_score
        a normalized RNA-seq RPKM object, which produced by "tsea.expression.normalization".

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 recommended "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 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 specificity. Bioinformatics, In submission.

See Also

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

Examples

```
#load("data/query_GTEx.rda")
#query_matrix = query_GTEx
#load("data/correction_factor.rda")
#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")
```

```
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, 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 specificity 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 specificity. Bioinformatics, In submission.

See Also

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

Examples

```
#load("data/query_GTEx.rda")
#query_matrix = query_GTEx
#load("data/correction_factor.rda")
#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")
```

| 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 specificity. Bioinformatics, In submission.

See Also

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

Examples

```
#load("data/GWAS_gene_multiple.rda")
#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")
#tsea.plot(tsea_t_multi, 0.05)
```

---

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 specificity. Bioinformatics, In submission.

See Also

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

Examples

```
#load("data/query_GTEx.rda")
#query_matrix = query_GTEx
#load("data/correction_factor.rda")
#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")
#tseaed_in_ENCODE_summary = tsea.summary(tseaed_in_ENCODE)
```

Index