

Package ‘TPAC’

October 26, 2023

Type Package

Title Tissue-Adjusted Pathway Analysis of Cancer (TPAC)

Version 0.2.0

Description Contains logic for single sample gene set testing of cancer transcriptomic data with adjustment for normal tissue-specificity.

Frost, H. Robert (2023) ``Tissue-adjusted pathway analysis of cancer (TPAC)" <doi:10.1101/2022.03.17.484779>.

Depends R (>= 3.6.0)

Imports data.table, MASS, TPACData

License GPL (>= 2)

Encoding UTF-8

NeedsCompilation no

Author H. Robert Frost [aut, cre]

Maintainer H. Robert Frost <rob.frost@dartmouth.edu>

Repository CRAN

Date/Publication 2023-10-26 15:20:08 UTC

R topics documented:

TPAC-package	2
createGeneSetCollection	2
getSupportedCancerTypes	3
tpac	4
tpacForCancer	5
tpacForCollection	7
Index	9

TPAC-package

Tissue-adjusted pathway analysis of cancer (TPAC)

Description

Implementation of TPAC, a method for single sample gene set testing of cancer transcriptomics data that leverages normal tissue-specificity.

Details

Package: TPAC
Type: Package
Version: 0.1.0
Date: 2023
License: GPL-2

Note

This work was supported by the National Institutes of Health grants R35GM146586, R21CA253408, P20GM130454 and P30CA023108.

Author(s)

H. Robert Frost

References

- Frost, H. R. (2022). Tissue-adjusted pathway analysis of cancer (TPAC). biorXiv e-prints. doi: <https://doi.org/10.1101/2022.03.17.484779>

createGeneSetCollection

Utility function to help create gene set collection list object

Description

Utility function that creates a gene set collection list in the format required by [tpacForCollection](#) given the gene IDs measured in the expression matrix and a list of gene sets as defined by the IDs of the member genes.

Usage

```
createGeneSetCollection(gene.ids, gene.set.collection, min.size=1, max.size)
```

Arguments

gene.ids	Vector of gene IDs. This should correspond to the genes measured in the gene expression data.
gene.set.collection	List of gene sets where each element in the list corresponds to a gene set and the list element is a vector of gene IDs. List names are gene set names. Must contain at least one gene set.
min.size	Minimum gene set size after filtering out genes not in the gene.ids vector. Gene sets whose post-filtering size is below this are removed from the final collection list. Default is 1 and cannot be set to less than 1.
max.size	Maximum gene set size after filtering out genes not in the gene.ids vector. Gene sets whose post-filtering size is above this are removed from the final collection list. If not specified, no filtering is performed.

Value

Version of the input gene.set.collection list where gene IDs have been replaced by position indices, genes not present in the gene.ids vector have been removed and gene sets failing the min/max size constraints have been removed.

See Also

[tpacForCollection](#)

Examples

```
# Create a collection with two sets defined over 3 genes
createGeneSetCollection(gene.ids=c("A", "B", "C"),
  gene.set.collection = list(set1=c("A", "B"), set2=c("B", "C")),
  min.size=2, max.size=3)
```

getSupportedCancerTypes

Get supported cancer types

Description

Retrieves the names of cancer types supported by the [tpacForCancer](#) function.

Usage

```
getSupportedCancerTypes()
```

Value

Vector containing the names of supported cancer types.

See Also

[tpacForCancer](#)

Examples

```
getSupportedCancerTypes()
```

 tpac

Tissue-adjusted pathway analysis of cancer (TPAC) algorithm

Description

Implementation of the tissue-adjusted pathway analysis of cancer (TPAC) method, which performs single sample gene set scoring of cancer transcriptomic data. The TPAC method computes a tissue-adjusted squared Mahalanobis distance and one-sided CDF values for all samples in the specified cancer gene expression matrix. Importantly, these distances are measured from the normal tissue mean and normal tissue specificity (i.e., fold-change in mean expression in the corresponding normal tissue relative to the mean expression in other tissue types) is used to weight genes when computing the distance metric. The distances are decomposed into a positive component, negative component and overall component, i.e., the positive component only captures expression values above the mean. A gamma distribution is fit on the squared tissue-adjusted Mahalanobis distances computed on data that is permuted to match to the null that the cancer expression data is uncorrelated with mean values matching those found in the associated normal tissue. This null gamma distribution is used to calculate the CDF values for the unpermuted distances.

Usage

```
tpac(gene.expr, mean.expr, tissue.specificity)
```

Arguments

gene.expr	An $n \times p$ matrix of gene expression values for n samples and p genes. This matrix should reflect the subset of the full expression profile that corresponds to a single gene set.
mean.expr	Vector of mean expression values for the p genes from which distances will be measured, i.e., mean expression in corresponding normal tissue. These mean expression values must be computed using the same normalization method used for tumor expression data in gene.expr
tissue.specificity	Optional vector of tissue-specificity values for the p genes, i.e, fold change between mean expression in the associated normal tissue and the average of mean expression values in other cancer-associated normal tissues. If specified, these specificity values are used to modify the diagonal elements of the covariance matrix used in the Mahalanobis statistic.

Value

A data.frame with the following elements (row names will match row names from gene.expr):

- s.pos: vector of TPAC scores (i.e., CDF values) computed from the positive squared adjusted Mahalanobis distances.
- s.neg: vector of TPAC scores computed from the negative squared adjusted Mahalanobis distances.
- s: vector of TPAC scores computed from the sum of the positive and negative squared adjusted Mahalanobis distances.

See Also

[tpacForCollection](#)

Examples

```
# Simulate Gaussian expression data for 10 genes and 10 samples
gene.expr=matrix(rnorm(100), nrow=10)
# Use 0 as mean.expr
mean.expr=rep(0,10)
# Simulate tissue-specific weights
tissue.specificity = runif(10, min=0.5, max=1.5)
# Execute TPAC to compute positive, negative and overall scores
tpac(gene.expr=gene.expr, mean.expr=mean.expr, tissue.specificity=tissue.specificity)
```

tpacForCancer

Executes the TPAC method on cancer gene expression data

Description

Executes the TPAC (tissue-adjusted pathway analysis for cancer) method ([tpacForCollection](#)) on cancer gene expression data using normal tissue expression data from the Human Protein Atlas (HPA) that is included in the package as [hpa.data](#). This HPA normal tissue data was specially processed by the HPA group as FPKM values using a pipeline similar to that employed by GDC for the TCGA data. For consistency with this HPA normal tissue data, the provided cancer.gene.expr data must be specified as FPKM+1 values. Please see the vignette for an example of calling this function using appropriately normalized TCGA gene expression data.

Usage

```
tpacForCancer(cancer.gene.expr, cancer.type, gene.set.collection,
              min.set.size=1, max.set.size)
```

Arguments

<code>cancer.gene.expr</code>	An $n \times p$ matrix of gene expression values for n tumors of the specified tumor type and p genes. The data should be normalized as FPKM+1 values, row names should be sample ID, and column names should be Ensembl gene IDs.
<code>cancer.type</code>	Cancer type of the expression data. Must be one of the supported cancer types as per getSupportedCancerTypes .
<code>gene.set.collection</code>	List of m gene sets for which scores are computed. Each element in the list corresponds to a gene set and the list element is a vector of Ensembl IDs for genes in the set. Gene set names should be specified as list names.
<code>min.set.size</code>	See description of <code>min.size</code> in createGeneSetCollection
<code>max.set.size</code>	See description of <code>max.size</code> in createGeneSetCollection

Value

A list containing two elements:

- `S.pos`: $n \times m$ matrix of TPAC scores computed using the positive squared adjusted Mahalanobis distances.
- `S.neg`: $n \times m$ matrix of TPAC scores computed using the negative squared adjusted Mahalanobis distances.
- `S`: $n \times m$ matrix of TPAC scores computed using the sum of the positive and negative squared adjusted Mahalanobis distances.

See Also

[tpac](#), [hpa.data](#), [tpacForCollection](#), [getSupportedCancerTypes](#), [createGeneSetCollection](#)

Examples

```
# Simulate Gaussian expression data for 10 genes and 10 samples
# (Note: cancer expression should be FPKM+1 for real applications)
cancer.gene.expr=matrix(rnorm(200), nrow=20)
# Create arbitrary Ensembl IDs
gene.ids = c("ENSG00000000003", "ENSG00000000005", "ENSG00000000419",
             "ENSG00000000457", "ENSG00000000460", "ENSG00000000938",
             "ENSG00000000971", "ENSG00000001036", "ENSG00000001084",
             "ENSG00000001167")
colnames(cancer.gene.expr) = gene.ids
# Define a collection with two disjoint sets that span the 10 genes
collection=list(set1=gene.ids[1:5], set2=gene.ids[6:10])
# Execute TPAC on both sets
tpacForCancer(cancer.gene.expr=cancer.gene.expr, cancer.type="glioma",
              gene.set.collection=collection)
```

tpacForCollection	<i>TPAC method for multiple gene sets</i>
-------------------	---

Description

Executes the TPAC (tissue-adjusted pathway analysis for cancer) method ([tpac](#)) on multiple gene sets, i.e., a gene set collection.

Usage

```
tpacForCollection(gene.expr, mean.expr, tissue.specificity, gene.set.collection)
```

Arguments

gene.expr See description in [tpac](#)

mean.expr See description in [tpac](#)

tissue.specificity

See description in [tpac](#)

gene.set.collection

List of m gene sets for which scores are computed. Each element in the list corresponds to a gene set and the list element is a vector of indices for the genes in the set. The index value is defined relative to the order of genes in the gene.expr matrix. Gene set names should be specified as list names.

Value

A list containing two elements:

- s.pos: n x m matrix of TPAC scores computed using the positive squared adjusted Mahalanobis distances.
- s.neg: n x m matrix of TPAC scores computed using the negative squared adjusted Mahalanobis distances.
- s: n x m matrix of TPAC scores computed using the sum of the positive and negative squared adjusted Mahalanobis distances.

See Also

[tpac](#), [tpacForCancer](#)

Examples

```
# Simulate Gaussian expression data for 10 genes and 10 samples
gene.expr=matrix(rnorm(100), nrow=10)
# Use 0 as mean.expr
mean.expr=rep(0,10)
# Simulate tissue-specific weights
tissue.specificity = runif(10, min=0.5, max=1.5)
```

```
# Define a collection with two disjoint sets that span the 10 genes
collection=list(set1=1:5, set2=6:10)
# Execute TPAC on both sets
tpacForCollection(gene.expr=gene.expr, mean.expr=mean.expr,
  tissue.specificity=tissue.specificity, gene.set.collection=collection)
```


Index

* **file**

- createGeneSetCollection, 2
- getSupportedCancerTypes, 3
- tpac, 4
- tpacForCancer, 5
- tpacForCollection, 7

* **package**

- TPAC-package, 2

createGeneSetCollection, 2, 6

getSupportedCancerTypes, 3, 6

hpa.data, 5, 6

tpac, 4, 6, 7

TPAC-package, 2

tpacForCancer, 3, 4, 5, 7

tpacForCollection, 2, 3, 5, 6, 7