Package 'TiMEx'

November 6, 2016

Type Package

Title Finding Mutually Exclusive Groups of Alterations in Cancer Datasets

Description The implementation of a waiting time model for mutually exclusive cancer alterations. The method can be readily applied to any large cancer dataset available as a binary matrix.

Version 0.99.0

Date 2016-12-1

Author Simona Cristea <scristea@jimmy.harvard.edu>

Maintainer Simona Cristea <scristea@jimmy.harvard.edu>

Depends R (>= 3.2.2)

Imports gtools (>= 3.5.0), graphics (>= 3.1.0), igraph (>= 1.0.1), RBGL (>= 1.50.0), stats (>= 3.1.0)

Suggests rmarkdown,knitr

VignetteBuilder rmarkdown

biocViews Software, StatisticalMethod, SystemsBiology, GraphAndNetwork, Clustering

License GPL (>=2)

URL https://github.com/cbg-ethz/TiMEx

RoxygenNote 5.0.1

R topics documented:

`TiMEx-package` *The main usages of TiMEx*

Description

The main usages of TiMEx, a package for finding mutually exclusive groups of alterations in large cancer datasets.

Overview

The most important function in this package is `TiMEx`, which identifies all mutually exclusive groups in a binary dataset. TiMEx is a procedure implementing three steps: first, all pairs in the input dataset are tested for mutual exclusivity. Second, maximal cliques are identified on the basis of a selected number of pairs. Third, the resulting cliques are tested for mutual exclusivity. Additional inputs to `TiMEx` include thresholds on the significance and intensity of the mutually exclusive pairs (`pairMu` and `pairPvalue`) and q-value cutoff on the mutually exclusive groups (`groupPvalue`). Unless otherwise specified, `TiMEx` will use the default values of these inputs.

Alternatively, users interested in running separately the three steps of the TiMEx procedure should run the functions `analyzePairs`, `doMaxCliques`, and `findSignifCliques` (sequentially and in this order).

Preprocessing and postprocessing

This package also provides functions to preprocess the input data (`doMetagene`, `removeLowFreqs`), as well as to postprocess the identified mutually exclusive groups (`produceTablesSignifGroups`, `subsampleAnalysis`, `plotGroupByName`, `recoverAllNamesGroups`).

Datasets

Multiple datasets are available within this package. `breast` and `ovarian` are datasets downloaded from cBioPortal (TCGA) in July 2014, and preprocessed as explained in "TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu et al. (Bioinformatics, 2016). `gbmDendrix` is a glioblastoma dataset used by Leiserson *et. al* in "Simultaneous identification of multiple driver pathways in cancer" (Plos Computational Biology, 2013). Additionally, the dataset `gbm`, preprocessed by Szczurek *et. al* as explained in "Modeling mutual exclusivity of cancer mutations" (Research in Computational Molecular Biology, 2014), and available in the package `muex` (`https://www1.ethz.ch/bsse/cbg/software/muex`), was also used in examples in this package.

For each of these four datasets, the identified significantly mutually exclusive groups are available as a separate dataset (`breastOutput`, `ovarianOutput`, `gbmDendrixOutput`, and `gbmMuexOutput`). Similarly, results of a subsampling analysis ran with 100 repetitions on the identified groups are available as separate datasets (`breastSubsampling`, `ovarianSubsampling`, `gbmDendrixSubsampling`, and `gbmMuexSubsampling`).

In the case of breast and ovarian cancer, the metagroups of genes in the original datasets (produced with the function `doMetagene`) are available as separate datasets (`breastGroups` and `ovarianGroups`).

Finally, the binary input matrices corresponding to the four breast cancer subtypes "LuminalA", "LuminalB", "Her2", and "Basal" are available in the dataset `breastSubtypes`, and the significantly mutually exclusive groups identified in each of these four subtypes are available in the dataset `breastSubtypesOutput`.

Simulations

Datasets can be generated from the TiMEx model using the function `simulateGenes`.

More

For more in-depth explanations of the TiMEx package and model, including examples, please see the corresponding vignette.

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016)

`analyzePairs` *Finds all mutually exclusive pairs in a dataset*

Description

`analyzePairs` performs step 1 of the TiMEx procedure: tests all gene pairs for mutual exclusivity. It returns a complex list, including parameter estimates, pairwise p-values and intensities of mutual exclusivity, log likelihoods, and others.

Usage

```
analyzePairs(mat)
```

Arguments

`mat`	binary alteration matrix, with rows representing patients and columns representing genes

Details

In the first step, the TiMEx procedure for identifying mutually exclusive groups of alterations tests all pairs for mutual exclusivity. The data corresponding to each pair of genes is therefore fitted to both the Null (Conditional Independence) and the Mutual Exclusivity models. Parameters under the two models are estimated, and, since they are nested, a likelihood ratio test is performed between the corresponding log likelihoods, in order to test whether mu (the intensity of mutual exclusivity) is different from 0. For more details on the TiMEx procedure, as well as on the underlying mathematical model, see "TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

The output list contains exhaustive information on the pairwise testing of genes, including parameter estimates, pairwise p-values and intensities of mutual exclusivity, log likelihoods, and others.

This function displays progress messages indicating the gene which is currently being tested against the remaining genes.

Value

List consisting of a set of matrices, all of dimension `n x n` (`n` being the number of genes). Element `(i,j)` of each matrix corresponds to the pairwise test between genes *i* and *j*:

- `lamiEstNull` matrix with rate estimates (lambda) of the waiting time of one gene (gene `i`) under the Null model.
- `lamjEstNull` matrix with rate estimates (lambda) of the waiting time of the other gene (gene `j`) under the Null model.
- `lamiEstME` matrix with rate estimates (lambda) of the waiting time of one gene (gene `i`) under the Mutual Exclusivity model.
- `lamjEstME` matrix with rate estimates (lambda) of the waiting time of the other gene (gene `j`) under the Mutual Exclusivity model.
- `likeNull` matrix with log likelihoods under the Null model.
- `likeAlt` matrix with log likelihoods under the Mutual Exclusivity model.
- `LRT` matrix with log likelihood ratio test (LRT) statistics.
- `pvalueLRTCorrectSym` list of the following symmetric matrices:
  - `bonferroni` bonferroni corrected p-values corresponding to the LRT statistic.
  - `bonferroni` fdr corrected p-values corresponding to the LRT statistic.
  - `uncorrected` uncorrected p-values corresponding to the LRT statistic.
- `muEstSym` symmetric matrix with estimated pairwise mu (intensity of mutual exclusivity)

Author(s)

Simona Cristea, `<scristea@jimmy.harvard.edu>`

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016)

See Also

doMaxCliques for step 2 of the TiMEx procedure; findSignifCliques for step 3 of the TiMEx procedure; the wrapper function TiMEx for combining these three steps, and identifying mutually exclusive groups in a binary dataset with the TiMEx model.

Examples

```
# Test all pairs from the ovarian dataset for mutual exclusivity (takes
# approximately 5 minutes)
data(ovarian)
ovarianPairs<-analyzePairs(ovarian)
```

---

breast *Breast cancer dataset*

Description

Dataset containing a binary alteration pattern for the breast cancer dataset downloaded from cBioPortal (TCGA) in July 2014, and preprocessed as explained in "TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016). Rows represent patients, and columns represent alterations.

Format

breast is a binary matrix with 958 rows and 537 columns.

Source

http://www.cbioportal.org/study.do?cancer_study_id=brca_tcga

---

breastGroups *Metagroups of genes in breast cancer*

Description

Dataset containing the genes with identical alteration patterns in the breast cancer dataset breast (before preprocessing). It is represented as a list of metagenes, with as many elements as input genes which had an identical alteration pattern with at least one other input gene.

Format

breastGroups is a list with 273 elements, where each element is a vector of genes with identical alteration patterns as the current gene. The numbers indicate the positions of the genes in the input matrix.

Source

Produced with the function doMetagene.

`breastOutput` *Mutually exclusive groups in breast cancer*

Description

Dataset containing the members of the groups identified as significantly mutually exclusive by TiMEx in breast cancer, together with their intensities of mutual exclusivity, corrected p-values, and other information.

Format

`breastOutput` is a list consisting of:

- `genesSignif` list of significantly mutually exclusive groups, as gene names, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, `genesSignif[[2]]` is a list containing the gene names of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent gene names of significantly mutually exclusive groups.
- `idxSignif` list of significantly mutually exclusive groups, as indices in the input matrix, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, `idxSignif[[2]]` is a list containing the indices of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent indices of significantly mutually exclusive groups.
- `pvals` list of corrected significant p-values corresponding to the tested cliques, ordered ascendingly. The list contains as many elements as identified lengths of significant groups. For example, `pvals[[2]]` is a list containing the p-values of the significant maximal cliques of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `posSignif` list of positions of the significant groups in the input list of maximal cliques, ordered ascendingly by corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, `posSignif[[2]]` is a list containing the positions of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `MusGroup` list of inferred mu values corresponding to the tested cliques, ordered ascendingly by the corresponding corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, `MusGroup[[2]]` is a list containing the mu values of the significant maximal cliques of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `mcStruct` input structure of maximal cliques to be tested for mutual exclusivity, as returned by `doMaxCliques`.
- `matrix` input binary alteration matrix.
- `groupPvalue` input threshold for the corrected p-value, lower than which cliques are significant.

Source

Produced with the function TiMEx, on the binary matrix in the input dataset breast.

---

breastSubsampling *Stability of mutually exclusive groups in breast cancer*

Description

Dataset containing the stability of the mutually exclusive groups identified by TiMEx in the breast cancer dataset breast, after subsampling the set of patients at frequencies of 30%, 50%, and 80%, for 100 times.

Format

breastSubsampling is a list with as many elements as subsampling frequencies provided (3 in this case). Each element is further a list with as many elemenets as number of sizes of the significantly mutually exclusive groups identified. Aditionally, bonf and fdr are two lists corresponding to each of these elements, representing different multiple correction methods. Finally, each element is a vector of relative counts of the significantly mutually exclusive groups identified. For example, breastSubsampling[[1]][[3]] represents the relative counts of the identified mutually exclusive groups of size 3 for a subsampling frequency of 30%, for both fdr and bonf (bonferroni) multiple correction methods.

Source

Produced with the function subsampleAnalysis, ran with the inputs subsampl<-c(0.3,0.5,0.8), noReps<-100, and the mutually exclusive groups from breastOutput.

---

breastSubtypes *Breast cancer subtypes*

Description

Dataset containing binary alteration patterns for the breast cancer subtypes "luminalA", "luminalB", "Her2", and "Basal2", downloaded from cBioPortal (TCGA) in July 2014, and preprocessed as explained in "TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016). Rows represent patients, and columns represent alterations.

Format

breastSubtypes is a list with 4 elements, corresponding to the 4 breast subtypes. Each element is binary matrix with 537 columns, as follows: breastSubtypes$luminalA consists of 222 rows, breastSubtypes$luminalB consists of 125 rows, Her2 consists of 55 rows, and Basal consists of 76 rows.

Source

http://www.cbioportal.org/study.do?cancer_study_id=brca_tcga

`breastSubtypesOutput` *Mutually exclusive groups in breast cancer subtypes*

Description

Dataset containing the members of the groups identified as significantly mutually exclusive by TiMEx in each of the 4 breast cancer subtypes "luminalA", "luminalB", "Her2", and "Basal", together with their corresponding intensities of mutual exclusivity, corrected p-values, and other information.

Format

`breastSubtypesOutput` is a list consisting of 4 lists, corresponding to the 4 breast subtypes. Each list further consists of:

- `genesSignif` list of significantly mutually exclusive groups, as gene names, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, `genesSignif[[2]]` is a list containing the gene names of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent gene names of significantly mutually exclusive groups.
- `idxSignif` list of significantly mutually exclusive groups, as indices in the input matrix, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, `idxSignif[[2]]` is a list containing the indices of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent indices of significantly mutually exclusive groups.
- `pvals` list of corrected significant p-values corresponding to the tested cliques, ordered ascendingly. The list contains as many elements as identified lengths of significant groups. For example, `pvals[[2]]` is a list containing the p-values of the significant maximal cliques of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `posSignif` list of positions of the significant groups in the input list of maximal cliques, ordered ascendingly by corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, `posSignif[[2]]` is a list containing the positions of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `MusGroup` list of inferred mu values corresponding to the tested cliques, ordered ascendingly by the corresponding corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, `MusGroup[[2]]` is a list containing the mu values of the significant maximal cliques of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `mcStruct` input structure of maximal cliques to be tested for mutual exclusivity, as returned by `doMaxCliques`.
- `matrix` input binary alteration matrix.
- `groupPvalue` input threshold for the corrected p-value, lower than which cliques are significant.

Source

Produced with the function TiMEx, on the four binary matrices in the input dataset breastSubtypes.

---

doMaxCliques *Identifies maximal cliques from pairwise testing information*

Description

doMaxCliques performs step 2 of the TiMEx procedure: identifies maximal cliques using information from pairwise testing. The maximal clique detection routine only uses the connections between gene pairs which satisfy the thresholds on mu (pairMu) and pvalue (pairPvalue).

Usage

```
doMaxCliques(pairs, pairMu, pairPvalue)
```

Arguments

pairs	list resulting after pairwise testing, as returned by analyzePairs
pairMu	pair-level threshold on mu (real number between 0 and 1). Default is 0.5.
pairPvalue	pair-level threshold on p-value (real number between 0 and 1). Default is 0.01.

Details

In the second step, the TiMEx procedure for identifying mutually exclusive groups of alterations detects maximal cliques using pairwise testing information from step 1. A graph is constructed, in which genes are vertices, and an edge is drawn between any pair *(i,j)* if both the estimated intensity of mutual exclusivity and the computed p-value satisfy the chosen thresholds pairMu and pairPvalue. Maximal cliques are detected on this graph.

The two thresholds can be set by the user, and are recommended to be chosen based on the sensitivity and specificity levels to which they correspond, as assessed in simulated data. For details, see "TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (2015). The default values are 0.5 for pairMu and 0.01 for pairPvalue.

This function needs functions from the packages *RBGL* and *igraph* to run.

Value

list consisting of:

- detectedLengths vector of lengths of the identified maximal cliques.
- idxInCliques list with as many elements as lengths of the identified maximal cliques. Each element of the list is a matrix, in which each row represents the indices of genes in an identified maximal clique.
- genesInCliques list with as many elements as lengths of the identified maximal cliques. Each element of the list is a matrix, in which each row represents the names of genes in an identified maximal clique.
- noMaxCliques vector of numbers of identified maximal cliques corresponding to each length present in the field detectedLengths.

- Mus list of two elements: OrderedGenesInCliques and OrderedIdxInCliques, which have the same structure as the elements genesInCliques and idxInCliques. The only difference is that the identified maximal cliques are now ordered by their average pairwise mu.
- pairMu input pair-level threshold on mu.
- pairPvalue input pair-level threshold on p-value.

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

analyzePairs for step 1 of the TiMEx procedure; findSignifCliques for step 3 of the TiMEx procedure; the wrapper function TiMEx for combining these three steps, and identifying mutually exclusive groups in a binary dataset with the TiMEx model.

Examples

```
# First, test all pairs from the ovarian cancer dataset for mutual
# exclusivity (take approximately 5 minutes)
data(ovarian)
ovarianPairs<-analyzePairs(ovarian)

# Then, identify all maximal cliques using the default thresholds
ovarianMaxCliques<-doMaxCliques(ovarianPairs)
```

---

doMetagene *Creates metagroups of genes*

Description

doMetagene collapses genes with identical alteration patterns across patients into metagroups. It returns a new matrix with the collapsed genes, as well as the members of the metagroups.

Usage

```
doMetagene(mat)
```

Arguments

mat	binary alteration matrix, with rows representing patients and columns representing genes

Details

It is recommended to run this function on the input binary matrix before applying TiMEx, because genes with identical alteration patterns across patients will otherwise be indistinguishable.

Note that in the datasets provided in this package, the genes have already been collapsed into metagroups.

Value

List consisting of:

- `newMat` the collapsed input binary matrix, with metagenes instead of genes.
- `groups` list of metagenes, with as many elements as input genes which had an identical alteration pattern with at least one other input gene.

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

`ovarianGroups`, `breastGroups` for examples of metagroups in large cancer datasets.

Examples

```
# Simulate genes and extract groups
simGenes<-simulateGenes(c(0.5,1,0.3),0.8,4000)
genes<-cbind(simGenes$genes,simGenes$genes)
genesNew<-doMetagene(genes)

# In the datasets provided in this package, the genes have already been
# collapsed into metagroups, hence the new matrix will be identical to the
# old one.
data(ovarian)
ovarianNew<-doMetagene(ovarian)
```

`findSignifCliques` *Tests all maximal cliques for mutual exclusivity*

Description

`findSignifCliques` performs step 3 of the TiMEx procedure, namely tests all candidate maximal cliques for mutual exclusivity and reports the significant ones.

Usage

```
findSignifCliques(mat, mcStruct, groupPvalue)
```

Arguments

mat	binary alteration matrix, with rows representing patients and columns representing genes
mcStruct	list containing maximal cliques, as returned by doMaxCliques
groupPvalue	threshold for the corrected p-value of the groups, lower than which cliques are significant (real number between 0 and 1). Default is 0.1.

Details

This function displays progress messages, namely the size of the clique currently being tested, and the number of cliques to test.

Note that sequentially performing steps 1, 2, and 3 of the TiMEx procedure (functions analyzePairs, doMaxCliques, and findSignifCliques) is equivalent to simply running the function TiMEx.

Value

list consisting of:

- genesSignif list of significantly mutually exclusive groups, as gene names, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, genesSignif[[2]] is a list containing the gene names of the significant groups of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent gene names of significantly mutually exclusive groups.
- idxSignif list of significantly mutually exclusive groups, as indices in the input matrix, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, idxSignif[[2]] is a list containing the indices of the significant groups of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent indices of significantly mutually exclusive groups.
- pvals list of corrected significant p-values corresponding to the tested cliques, ordered ascendingly. The list contains as many elements as identified lengths of significant groups. For example, pvals[[2]] is a list containing the p-values of the significant maximal cliques of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- posSignif list of positions of the significant groups in the input list of maximal cliques, ordered ascendingly by corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, posSignif[[2]] is a list containing the positions of the significant groups of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- MusGroup list of inferred mu values corresponding to the tested cliques, ordered ascendingly by the corresponding corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, MusGroup[[2]] is a list containing the mu values of the significant maximal cliques of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- mcStruct input structure of maximal cliques to be tested for mutual exclusivity, as returned by doMaxCliques

- matrix input binary alteration matrix
- groupPvalue input threshold for the corrected p-value, lower than which cliques are significant

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

analyzePairs for step 1 of the TiMEx procedure; doMaxCliques for step 2 of the TiMEx procedure; the wrapper function TiMEx for combining these three steps, and identifying mutually exclusive groups in a binary dataset with the TiMEx model. The data structures ovarianOutput, breastOutput, gbmDendrixOutput, and gbmMuexOutput are examples of structures resulting after running TiMEx on large cancer datasets.

Examples

```
# First, test all pairs from the ovarian dataset for mutual exclusivity
# (takes approximately 5 minutes).
data(ovarian)
ovarianPairs<-analyzePairs(ovarian)

# Second, identify all maximal cliques using the default thresholds
ovarianMaxCliques<-doMaxCliques(ovarianPairs)

# Then, test all maximal cliques for mutual exclusivity and report the
# significant ones, based on a corrected p-value threshold of 0.1 (default).
ovarianMEgroups<-findSignifCliques(ovarian,ovarianMaxCliques)
```

---

gbmDendrix *Glioblastoma dataset used by Dendrix*

Description

Dataset containing a binary alteration pattern for the glioblastoma dataset used by Leiserson *et. al* in "Simultaneous identification of multiple driver pathways in cancer" (Plos Computational Biology, 2013). Rows represent patients, and columns represent alterations. In the names of alterations, *(D)* represents a copy number deletion, and *(A)* represents a copy number amplification.

Format

gbmDendrix is a binary matrix with 261 rows and 486 columns.

Source

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003054

gbmDendrixOutput *Mutually exclusive groups in the glioblastoma dataset used by Dendrix*

Description

Dataset containing the members of the groups identified as significantly mutually exclusive by TiMEx in the glioblastoma dataset used by Leiserson *et. al* in "Simultaneous identification of multiple driver pathways in cancer" (Plos Computational Biology, 2013), together with their intensities of mutual exclusivity, corrected p-values, and other information.

Format

`gbmDendrixOutput` is a list consisting of:

- `genesSignif` list of significantly mutually exclusive groups, as gene names, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, `genesSignif[[2]]` is a list containing the gene names of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent gene names of significantly mutually exclusive groups.
- `idxSignif` list of significantly mutually exclusive groups, as indices in the input matrix, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, `idxSignif[[2]]` is a list containing the indices of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent indices of significantly mutually exclusive groups.
- `pvals` list of corrected significant p-values corresponding to the tested cliques, ordered ascendingly. The list contains as many elements as identified lengths of significant groups. For example, `pvals[[2]]` is a list containing the p-values of the significant maximal cliques of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `posSignif` list of positions of the significant groups in the input list of maximal cliques, ordered ascendingly by corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, `posSignif[[2]]` is a list containing the positions of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `MusGroup` list of inferred mu values corresponding to the tested cliques, ordered ascendingly by the corresponding corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, `MusGroup[[2]]` is a list containing the mu values of the significant maximal cliques of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `mcStruct` input structure of maximal cliques to be tested for mutual exclusivity, as returned by `doMaxCliques`.
- `matrix` input binary alteration matrix.
- `groupPvalue` input threshold for the corrected p-value, lower than which cliques are significant.

Source

Produced with the function TiMEx, on the binary matrix in the input dataset gbmDendrix.

---

gbmDendrixSubsampling *Stability of mutually exclusive groups in the glioblastoma datased used by Dendrix*

Description

Dataset containing the stability of the mutually exclusive groups identified by TiMEx in the glioblastoma dataset gbmDendrix, used by Leiserson *et. al* in "Simultaneous identification of multiple driver pathways in cancer" (Plos Computational Biology, 2013), after subsampling the set of patients at frequencies of 30%, 50%, and 80%, for 100 times.

Format

gbmDendrixSubsampling is a list with as many elements as subsampling frequencies provided (3 in this case). Each element is further a list with as many elemenets as number of sizes of the significantly mutually exclusive groups identified. Aditionally, bonf and fdr are two lists corresponding to each of these elements, representing different multiple correction methods. Finally, each element is a vector of relative counts of the significantly mutually exclusive groups identified. For example, gbmDendrixSubsampling[[1]][[3]] represents the relative counts of the identified mutually exclusive groups of size 3 for a subsampling frequency of 30%, for both fdr and bonf (bonferroni) multiple correction methods.

Source

Produced with the function subsampleAnalysis, ran with the inputs subsampl<-c(0.3,0.5,0.8), noReps<-100, and the mutually exclusive groups from gbmDendrixOutput.

---

gbmMuexOutput *Mutually exclusive groups in the glioblastoma dataset used by muex*

Description

Dataset containing the members of the groups identified as significantly mutually exclusive by TiMEx in the glioblastoma dataset used by Szczurek *et. al* in "Modeling mutual exclusivity of cancer mutations" (Research in Computational Molecular Biology, 2014), together with their intensities of mutual exclusivity, corrected p-values, and other information. The dataset which was used as input for producing these groups can be accessed via data(gbm) in the R package muex, available at https://www1.ethz.ch/bsse/cbg/software/muex.

Format

`gbmMuexOutput` is a list consisting of:

- `genesSignif` list of significantly mutually exclusive groups, as gene names, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, `genesSignif[[2]]` is a list containing the gene names of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent gene names of significantly mutually exclusive groups.
- `idxSignif` list of significantly mutually exclusive groups, as indices in the input matrix, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, `idxSignif[[2]]` is a list containing the indices of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent indices of significantly mutually exclusive groups.
- `pvals` list of corrected significant p-values corresponding to the tested cliques, ordered ascendingly. The list contains as many elements as identified lengths of significant groups. For example, `pvals[[2]]` is a list containing the p-values of the significant maximal cliques of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `posSignif` list of positions of the significant groups in the input list of maximal cliques, ordered ascendingly by corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, `posSignif[[2]]` is a list containing the positions of the significant groups of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `MusGroup` list of inferred mu values corresponding to the tested cliques, ordered ascendingly by the corresponding corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, `MusGroup[[2]]` is a list containing the mu values of the significant maximal cliques of size 2. Each list of this type further has two elements, `fdr` and `bonf`, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- `mcStruct` input structure of maximal cliques to be tested for mutual exclusivity, as returned by `doMaxCliques`.
- `matrix` input binary alteration matrix.
- `groupPvalue` input threshold for the corrected p-value, lower than which cliques are significant.

Source

Produced with the function `TiMEx`, on the binary matrix which can be accessed via `data(gbm)` in the R package `muex`.

gbmMuexSubsampling — *Stability of mutually exclusive groups in the glioblastoma datased used by muex*

Description

Dataset containing the stability of the mutually exclusive groups identified by TiMEx in the glioblastoma dataset used by Szczurek *et. al* in "Modeling mutual exclusivity of cancer mutations" (Research in Computational Molecular Biology, 2014), after subsampling the set of patients at frequencies of 30%, 50%, and 80%, for 100 times.

Format

`gbmMuexSubsampling` is a list with as many elements as subsampling frequencies provided. Each element is further a list with as many elemenets as number of sizes of the significantly mutually exclusive groups identified. Aditionally, `bonf` and `fdr` are two lists corresponding to each of these elements, representing different multiple correction methods. Finally, each element is a vector of subsampling frequencies of the significant mutually exclusive groups identified. For example, `gbmMuexSubsampling[[1]][[3]]` represents the relative counts of the identified mutually exclusive groups of size 3 for a subsampling frequency of 30%, for both `fdr` and `bonf` (bonferroni) multiple correction methods.

Source

Produced with the function `subsampleAnalysis`, ran with the inputs `subsampl<-c(0.3,0.5,0.8)`, `noReps<-100`, and the mutually exclusive groups from `gbmMuexOutput`.

ovarian — *Ovarian cancer dataset*

Description

Dataset containing a binary alteration pattern for the ovarian cancer dataset downloaded from cBioPortal (TCGA) in July 2014, and preprocessed as explained in "TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016). Rows represent patients, and columns represent alterations.

Format

A binary matrix with 316 rows and 312 columns.

Source

http://www.cbioportal.org/study.do?cancer_study_id=ov_tcga_pub

ovarianGroups *Metagroups of genes in ovarian cancer*

Description

Dataset containing the genes with identical alteration patterns in the ovarian cancer dataset ovarian (before preprocessing). It is represented as a list of metagenes, with as many elements as input genes which had an identical alteration pattern with at least one other input gene.

Format

ovarianGroups is a list with 263 elements, where each element is a vector of genes with identical alteration patterns as the current gene. The numbers indicate the positions of the genes in the input matrix.

Source

Produced with the function doMetagene.

ovarianOutput *Mutually exclusive groups in ovarian cancer*

Description

Dataset containing the members of the groups identified as significantly mutually exclusive by TiMEx in ovarian cancer, together with their intensities of mutual exclusivity, corrected p-values, and other information.

Format

ovarianOutput is a list consisting of:

- genesSignif list of significantly mutually exclusive groups, as gene names, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, genesSignif[[2]] is a list containing the gene names of the significant groups of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent gene names of significantly mutually exclusive groups.
- idxSignif list of significantly mutually exclusive groups, as indices in the input matrix, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, idxSignif[[2]] is a list containing the indices of the significant groups of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent indices of significantly mutually exclusive groups.
- pvals list of corrected significant p-values corresponding to the tested cliques, ordered ascendingly. The list contains as many elements as identified lengths of significant groups. For example, pvals[[2]] is a list containing the p-values of the significant maximal cliques of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.

- posSignif list of positions of the significant groups in the input list of maximal cliques, ordered ascendingly by corrected p-value. The list contains as many elements as identified lengths of significant groups.For example, posSignif[[2]] is a list containing the positions of the significant groups of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- MusGroup list of inferred mu values corresponding to the tested cliques, ordered ascendingly by the corresponding corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, MusGroup[[2]] is a list containing the mu values of the significant maximal cliques of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- mcStruct input structure of maximal cliques to be tested for mutual exclusivity, as returned by doMaxCliques.
- matrix input binary alteration matrix.
- groupPvalue input threshold for the corrected p-value, lower than which cliques are significant.

Source

Produced with the function TiMEx, on the binary matrix in the input dataset ovarian.

---

ovarianSubsampling *Stability of mutually exclusive groups in ovarian cancer*

Description

Dataset containing the stability of the mutually exclusive groups identified by TiMEx in the ovarian cancer dataset ovarian, after subsampling the set of patients at frequencies of 30%, 50%, and 80%, for 100 times.

Format

ovarianSubsampling is a list with as many elements as subsampling frequencies provided (3 in this case). Each element is further a list with as many elemenets as number of sizes of the significantly mutually exclusive groups identified. Aditionally, bonf and fdr are two lists corresponding to each of these elements, representing different multiple correction methods. Finally, each element is a vector of relative counts of the significantly mutually exclusive groups identified. For example, ovarianSubsampling[[1]][[3]] represents the relative counts of the identified mutually exclusive groups of size 3 for a subsampling frequency of 30%, for both fdr and bonf (bonferroni) multiple correction methods.

Source

Produced with the function subsampleAnalysis, ran with the inputs subsampl<-c(0.3,0.5,0.8), noReps<-100, and the mutually exclusive groups from ovarianOutput.

plotGroupByName *Plots a Mutually Exclusive group*

Description

plotGroupByName plots a mutually exclusive group (including the frequencies of the genes), given the names of the genes in the group.

Usage

```
plotGroupByName(group, mat)
```

Arguments

group	vector of gene names to be plotted
mat	binary alteration matrix, with rows representing patients and columns representing genes

Details

The plotting is done based on the function image.

Value

None

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

the wrapper function TiMEx for identifying mutually exclusive groups in a binary dataset with the TiMEx model.

Examples

```
# Plot the group consisting of the copy number aberrations of MIEN1 and
# CDKN1B, and the point mutations of CDH1, GATA3, and MAP3K1, in breast
# cancer.
data(breast)
group<-c("MIEN1-CNA","CDH1-Mut","GATA3-Mut","MAP3K1-Mut","CDKN1B-CNA")
plotGroupByName(group,breast)
```

produceTablesSignifGroups
Produces tables with groups

Description

produceTablesSignifGroups produces tables with the significant groups. These tables include the names of the genes part of the groups, their respective frequency in the dataset, and the mu and corrected pvalue corresponding to each group.

Usage

```
produceTablesSignifGroups(signifGroups, mat, noToShow)
```

Arguments

signifGroups	result structure with the significant groups, as returned by either TiMEx or findSignifCliques
mat	binary alteration matrix, with rows representing patients and columns representing genes
noToShow	maximum number of groups to include in the table. Default is 30.

Details

This function summarizes information on the mutually exclusive groups identified by TiMEx in a dataset, as tables.

Value

list with as many elements as lengths of the identified mutually exclusive groups, containing tables with the significant groups for each size. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent significantly mutually exclusive groups.

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

the wrapper function TiMEx for identifying mutually exclusive groups in a binary dataset with the TiMEx model.

Examples

```
# Produce tables on the output of TiMEx on the ovarian cancer dataset
data(ovarian)
data(ovarianOutput)
ovarianTables<-produceTablesSignifGroups(ovarianOutput,ovarian)
```

`recoverAllNamesGroups` *Recovers members of the metagroups*

Description

`recoverAllNameGroups` recovers, from a metagroup, the names of the genes part of an identified mutually exclusive group.

Usage

```
recoverAllNamesGroups(groupsMeta, clGenes)
```

Arguments

`groupsMeta`	list containing groups of equivalent genes, as returned by the field `groups` of `doMetagene`
`clGenes`	matrix of mutually exclusive groups of same size, as gene names. This type of matrix is returned by either `TiMEx` or `findSignifCliques`, as one of the matrix elements of the `genesSignif` field.

Details

This function can be used if the input binary matrix contains identical events that need to be merged into metagenes using `doMetagene`. Running `recoverAllNamesGroups` provides the set of identical alterations which are part of the identified mutually exclusive groups.

In order to run this function on all the identified mutually exclusive groups as returned by `TiMEx` or `findSignifCliques`, it is necessary to run it separately on each matrix element (corresponding to different group sizes and different correction methods) of the `genesSignif` field in the structure returned by either `TiMEx` or `findSignifCliques`.

For example, after loading `data(ovarianGroups)` and `data(ovarianOutput)`, and running

```
rGroups<-recoverAllNamesGroups(ovarianGroups,signifGroups$genesSignif[[3]]$bonf)
```

`rGroups[[14]]` has 3 elements (as many as genes part of the identified mutually exclusive groups). Each element is the metagroup of each of the genes part of the 14th mutually exclusive group in the input matrix `signifGroups$genesSignif[[3]]$bonf`. Namely, *BRD4-CNA* and *MYC-CNA* have unique alteration patterns among samples, and are alone in their metagroup, while *CASC1-CNA* has an identical alteration pattern with *KRAS-CNA* and *LYRM5-CNA*. The numbers below the gene names are the indices of the genes in the initial input binary matrix of patients.

Value

list with as many elements as number of identified mutually exclusive groups, *i.e.* number of rows in the input matrix. Each of its elements is further a list, containing, at each position, the metagroup of the genes in the initial group at that resepective position. For an example, see *Details* above.

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

doMetagene for collapsing the genes of an input matrix with identical alteration patterns into meta-groups.

Examples

```
data(ovarianGroups)
data(ovarianOutput)
r<-recoverAllNamesGroups(ovarianGroups,ovarianOutput$genesSignif[[3]]$bonf)
```

---

removeLowFreqs *Removes alterations based on frequency*

Description

removeLowFreqs returns a binary matrix from which genes altered with lower frequency than the input level are removed.

Usage

```
removeLowFreqs(mat, level)
```

Arguments

mat	binary alteration matrix, with rows representing patients and columns representing genes
level	frequency level under which the genes are to be removed (real number between 0 and 1). Default is 0.03.

Details

It is only recommended to run this function if the user is not at all interested in the role of low-frequently altered genes.

Value

the binary input matrix without the low-frequently altered genes.

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

`ovarian`, `breast`, and `gbmDendrix` for examples of biological large cancer datasets.

Examples

```
# Remove genes altered in less than 3% (the default level) of the samples
# in the ovarian cancer dataset
data(ovarian)
ovarianNew<-removeLowFreqs(ovarian)
```

`simulateGenes` *Generates data from the TiMEx model*

Description

`simulateGenes` returns a list containing a binary matrix simulated from the TiMEx model, for given lambdas (exponential rates), mu (intensity of mutual exclusivity), and N (sample size).

Usage

```
simulateGenes(lambdas, mu, N)
```

Arguments

`lambdas`	vector of exponential rates (positive real numbers). The length of the vector equals the number of simulated genes.
`mu`	intensity of mutual exclusivity (real number between 0 and 1). Default is 1 (perfect mutual exclusivity).
`N`	sample size (positive integer).

Details

This function needs `permutations` in order to run.

For details on how the values of the exponential rates correspond to frequencies, see *References* below.

Value

list consisting of:

- `genes` the simulated dataset, as a binary matrix.

- genoMat the matrix with genotype probabilities, from which the dataset was simulated. This matrix has as many dimensions as number of genes, i.e. 2x2x...x2. For each dimension, the first position corresponds to the probability of observing a 0 for that gene, and the second position corresponds to the probability of observing an 1. For example, in the case of 4 genes, the probability of observing the null genotype (0000) is given by genoMat[1,1,1,1]; the probability of observing the genotype (1011) is given by genoMat[2,1,2,2]; the probability of observing the genotype (1111) is given by genoMat[2,2,2,2]. The entries of this matrix are nonnegative and sum up to 1.

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

the wrapper function TiMEx for identifying mutually exclusive groups in a binary dataset with the TiMEx model, ovarian, breast, and gbmDendrix for examples of biological large cancer datasets.

Examples

```
simGenes<-simulateGenes(c(0.5,1,0.3),0.8,4000)
```

---

subsampleAnalysis *Assesses the stability of groups by subsampling*

Description

subsampleAnalysis subsamples the set of patients and assess the stability of the identified mutually exclusive groups.

Usage

```
subsampleAnalysis(subsampl, noReps, signifGroups)
```

Arguments

subsampl	a vector with subsampling frequencies
noReps	number of repetitions of subsampling. Default is 100.
signifGroups	result structure with the significant groups, as returned by either TiMEx or findSignifCliques

Details

As this function runs TiMEx many times sequentially, it is computationally very intensive. For a version of this function which can be directly ran on a cluster, please contact me (see e-mail below).

For example, after loading `data(ovarianOutput)` and running

`counts<-subsampleAnalysis(subsampl=c(0.3,0.5,0.8),signifGroups=signifGroups)`

`counts[[1]][[3]]` will represent the relative counts of the identified mutually exclusive groups of size 3 for a subsampling frequency of 30%, for both `fdr` and `bonf` (bonferroni) multiple correction methods.

Value

list with as many elements as subsampling frequencies provided. Each element is further a list with as many elemenets as number of sizes of the significantly mutually exclusive groups identified. Aditionally, `bonf` and `fdr` are two lists corresponding to each of these elements, representing different multiple correction methods. Finally, each element is a vector of subsampling frequencies of the significantly mutually exclusive groups identified. For an example, see *Details* above.

Author(s)

Simona Cristea, `<scristea@jimmy.harvard.edu>`

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

the wrapper function `TiMEx` for identifying mutually exclusive groups in a binary dataset with the TiMEx model, `ovarianSubsampling`, `breastSubsampling`, `gbmDendrixSubsampling`, and `gbmMuexSubsampling` for examples of outputs after performing the subsampling analysis.

Examples

```
# running this function is time-intensive
data(ovarianOutput)
## Not run: subsampleOvarian<-subsampleAnalysis(c(0.3,0.5,0.8),ovarianOutput)
```

---

testCliqueAsGroup *Tests whether a given group is mutually exclusive*

Description

`testCliqueAsGroup` tests whether a group, given as gene indices, is mutually exclusive.

Usage

```
testCliqueAsGroup(geneIdx, mat, lo)
```

Arguments

geneIdx	vector of indices in the input matrix of the genes to be tested
mat	binary alteration matrix, with rows representing patients and columns representing genes
lo	rate of observation time. Default is 1.

Details

For deciding whether a group is mutually exclusive, the group of genes is fitted to both the Null (Conditional Independence) and the Mutual Exclusivity models. Parameters under the two models are estimated, and, since they are nested, a likelihood ratio test is performed between the corresponding log likelihoods, in order to test whether mu (the intensity of mutual exclusivity) is different from 0. For computing the likelihood of the data under both models, an exhaustive enumeration of all possible orders of the input alterations needs to be performed. Therefore, the complexity of the test is exponential in the number of genes to be tested, which makes it unfeasible for large number of genes (usually more than 10).

lo (the rate of observation time) is by default set to 1, as both models are otherwise unindentifiable. We recommend leaving the value of this parameter unchanged, as otherwise the estimated waiting time rates of the genes require additional interpretation.

For more details on the TiMEx procedure, as well as on the underlying mathematical model, see "TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

Value

List consisting of:

- opMu list as returned by optim. The field par is a vector of n+1 positions (n being the number of genes) containing the estimates of the waiting time rates (lambda) for the n genes under the mutual exclusivity model, followed by the estimate for mu.
- opNull list as returned by optim. The field par is a vector of n positions (n being the number of genes) containing the estimates of the waiting time rates (lambda) for the n genes under the null model.
- countsVec the contingency table of the input genes, as a vector. The first element is the count of the samples where no gene was altered, the next n elements are the counts of the samples where exactly one gene was altered (starting with gene 1 and ending with gene n), the next n elements are the counts of the samples where exactly two genes were altered (starting with genes 1 and 2, continuing with genes 1 and 3, and ending with genes n-1 and n), and so on.
- genes subset of the input binary matrix, corresponding to the genes to be tested
- LRT log likelihood ratio (LRT)
- pvalueLRT the p-value corresponding to the LRT

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

the wrapper function TiMEx for identifying mutually exclusive groups in a binary dataset with the TiMEx model.

Examples

```
# Tests for mutual exclusivity the group of genes with indices 13, 204, and
# 310 in the ovarian cancer dataset
data(ovarian)
testGroup<-testCliqueAsGroup(c(13,204,310),ovarian)
```

TiMEx *Finds mutually exclusive groups*

Description

TiMEx is the main function of this package. It identifies all groups of mutually exclusive cancer alterations in a large binary input dataset.

Usage

```
TiMEx(mat, pairMu, pairPvalue, groupPvalue)
```

Arguments

mat	binary alteration matrix, with rows representing patients and columns representing genes
pairMu	pair-level threshold on mu (real number between 0 and 1). Default is 0.5.
pairPvalue	pair-level threshold on p-value (real number between 0 and 1). Default is 0.01.
groupPvalue	threshold for the corrected p-value, lower than which cliques are significant. Default to 0.1

Details

Dependning on the size of the dataset (both in terms of samples and alterations), TiMEx can require a reasonable time to run. For example, the approximate running time is 10 minutes for the ovarian cancer dataset, and 45 minutes for the breast cancer dataset included in this package, on a personal computer.

TiMEx displays progress messages. In a fist step, it indicates the gene which is currently being tested against the remaining genes. In a later step, it indicates the size of the clique currently being tested, and the number of cliques to test.

Value

list consisting of:

- genesSignif list of significantly mutually exclusive groups, as gene names, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, genesSignif[[2]] is a list containing the gene names of the significant groups of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent gene names of significantly mutually exclusive groups.
- idxSignif list of significantly mutually exclusive groups, as indices in the input matrix, sorted by corrected p-value. The list contains as many elements as identified lengths of groups. For example, idxSignif[[2]] is a list containing the indices of the significant groups of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a matrix, in which rows represent indices of significantly mutually exclusive groups.
- pvals list of corrected significant p-values corresponding to the tested cliques, ordered ascendingly. The list contains as many elements as identified lengths of significant groups. For example, pvals[[2]] is a list containing the p-values of the significant maximal cliques of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- posSignif list of positions of the significant groups in the input list of maximal cliques, ordered ascendingly by corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, posSignif[[2]] is a list containing the positions of the significant groups of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- MusGroup list of inferred mu values corresponding to the tested cliques, ordered ascendingly by the corresponding corrected p-value. The list contains as many elements as identified lengths of significant groups. For example, MusGroup[[2]] is a list containing the mu values of the significant maximal cliques of size 2. Each list of this type further has two elements, fdr and bonf, corresponding to different multiple testing correction methods. Each element is a vector, of length the number of significant maximal cliques of a given size.
- mcStruct input structure of maximal cliques to be tested for mutual exclusivity, as returned by doMaxCliques.
- matrix input binary alteration matrix.
- groupPvalue input threshold for the corrected p-value, lower than which cliques are significant.

Author(s)

Simona Cristea, <scristea@jimmy.harvard.edu>

References

"TiMEx: A Waiting Time Model For Mutually Exclusive Cancer Alterations", by Constantinescu *et al.* (Bioinformatics, 2016).

See Also

analyzePairs for step 1 of the TiMEx procedure; doMaxCliques for step 2 of the TiMEx procedure, and findSignifCliques for step 3 of the TiMEx procedure. The data structures ovarianOutput, breastOutput, gbmDendrixOutput, and gbmMuexOutput are examples of structures resulting after running TiMEx on large cancer datasets.

Examples

```
# Run TiMEx on the ovarian cancer dataset with default parameters
# (takes approximately 10 minutes)
data(ovarian)
ovarianMEGroups<-TiMEx(ovarian)
```

Index