DawnRank R Package Documentation

December 11, 2013

R topics documented:

DawnRank-package   *DawnRank*

Description

DawnRank calculates the importance of potential driver mutations in individual patientes using a pathway approach inspired by the PageRank algorithm. Mutations in genes that are more highly connected and connect to differentially expressed downstream genes are more likely to be driver mutations. DawnRank can calculate the importance of candidate driver mutations in a single patient as well as make predictions on the common drivers from a patient population using Condorcet Rank Aggregation.

Details

Package:	DawnRank
Type:	Package
Version:	1.1
Date:	2013-10-15
License:	GPL

Author(s)

Jack Hou <jackhou2@illinois.edu>, Jian Ma

Examples

```
###using a small subset of the TCGA dataset and a small KEGG
###gene interaction network,
###We will show how to get DawnRank Results

library(DawnRank)

#load the mutation data
data(brcaExampleMutation)

#load the tumor expression data
data(brcaExampleTumorExpression)

#load the normal expression data
data(brcaExampleNormalExpression)

#load the pathway data
data(brcaExamplePathway)

#load the gold standard
data(goldStandard)

#normalize the tumor and normal data to get the differential expression
normalizedDawn<-DawnNormalize(tumorMat=brcaExampleTumorExpression,
normalMat=brcaExampleNormalExpression)

#get the DawnRank Score
dawnRankScore<-DawnRank(adjMatrix=brcaExamplePathway,
mutationMatrix=brcaExampleMutation,expressionMatrix=normalizedDawn,
mu=3,goldStandard=goldStandard)

#look at the DawnRank scores for a few patients
dawnRankFrame<-dawnRankScore[[3]]
head(dawnRankFrame)

#get the aggregate DawnRank scores
aggregateDawnRankScore<-condorcetRanking(scoreMatrix=dawnRankScore[[2]],
mutationMatrix=brcaExampleMutation)

#look at top 10 ranked genes
top10<-aggregateDawnRankScore[[2]][1:10]
top10

#get the individual cutoff for patient TCGA-A2-A04P
dawnRankFrame$isCGC<-dawnRankFrame$isGoldStandard
library(maxstat)
significance<-patspeccutoff(patient="TCGA-A2-A04P",ms=dawnRankFrame,
default=95)
```

brcaExampleMutation   *brcaExampleMutation*

Description

Mutation data for a Breast Cancer Dataset from The Cancer Gene Atlas (TCGA). The data is presented in matrix with genes being the rows and patients being columns. A binary 0 or 1 is used to indicate whether or not a gene is mutated in that patient.

Usage

```
data(brcaExampleMutation)
```

Format

The format is: num [1:1492, 1:533] 0 0 0 0 0 0 0 0 0 0 ... - attr(*, "dimnames")=List of 2 ..$: chr [1:1492] "PML" "FGF3" "HLA-A" "CACNA1F" ... ..$: chr [1:533] "TCGA-A1-A0SD" "TCGA-A1-A0SE" "TCGA-A1-A0SH" "TCGA-A1-A0SJ" ...

Details

Mutation data for an example Breast Cancer Dataset. Rows are genes, columns are patients

Source

https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm

References

The Cancer Genome Atlas. Comprehensive molecular portraits of human breast tumors. Nature. 490 (7418):61-70.

Examples

```
data(brcaExampleMutation)
## maybe str(brcaExampleMutation) ; plot(brcaExampleMutation) ...
```

brcaExampleNormalExpression
                    brcaExampleNormalExpression

Description

Normal Expression data for an example Breast Cancer Dataset from TCGA. The data is in matrix with genes being the rows and patients being columns. This dataset contains the unnormalized normal expression of normal BRCA TCGA samples.

Usage

```
data(brcaExampleNormalExpression)
```

Format

The format is: num [1:1492, 1:62] 0.24 0.122 0.56 -0.231 1.189 ... - attr(*, "dimnames")=List of 2 ..$: chr [1:1492] "PML" "FGF3" "HLA-A" "CACNA1F" ... ..$: chr [1:62] "TCGA-A7-A0CE" "TCGA-A7-A0CH" "TCGA-A7-A0D9" "TCGA-A7-A0DB" ...

Details

Expression data for an example Breast Cancer Dataset. Rows are genes, columns are patients

Source

https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm

References

The Cancer Genome Atlas. Comprehensive molecular portraits of human breast tumors. Nature. 490 (7418):61-70.

Examples

```
data(brcaExampleNormalExpression)
## maybe str(brcaExampleNormalExpression) ;
```

---

brcaExamplePathway   *brcaExamplePathway*

Description

Pathway data (from KEGG) for an example Breast Cancer Dataset representing all networks in the to the KEGG "Pathways in Cancer". The pathway is in the form of an adjacency matrix describing curated gene-gene interactions.

Usage

```
data(brcaExamplePathway)
```

Format

The format is: num [1:1492, 1:1492] 0 0 0 0 0 0 0 0 0 0 ... - attr(*, "dimnames")=List of 2 ..$: chr [1:1492] "PML" "FGF3" "HLA-A" "CACNA1F" ... ..$: chr [1:1492] "PML" "FGF3" "HLA-A" "CACNA1F" ...

Details

An adjacency matrix detailing interactions of an example pathway

Source

http://www.genome.jp/kegg/

References

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M.; KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Res. 40, D109-D114 (2012).

Examples

```
data(brcaExamplePathway)
## maybe str(brcaExamplePathway) ; plot(brcaExamplePathway)...
```

```
brcaExampleTumorExpression
```
brcaExampleTumorExpression

Description

Tumor Expression data for an example Breast Cancer Dataset from the TCGA.The data is presented in matrix with genes being the rows and patients being columns. This dataset contains the unnormalized tumor expression.

Usage

```
data(brcaExampleTumorExpression)
```

Format

The format is: num [1:1492, 1:533] 0.0578 0.3625 0.7008 0.3775 0.4605 ... - attr(*, "dimnames")=List of 2 ..$: chr [1:1492] "PML" "FGF3" "HLA-A" "CACNA1F" ... ..$: chr [1:533] "TCGA-A1-A0SD" "TCGA-A1-A0SE" "TCGA-A1-A0SH" "TCGA-A1-A0SJ" ...

Details

Expression data for an example Breast Cancer Dataset. Rows are genes, columns are patients

Source

https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm

References

The Cancer Genome Atlas. Comprehensive molecular portraits of human breast tumors. Nature. 490 (7418):61-70.

Examples

```
data(brcaExampleTumorExpression)
## maybe str(brcaExampleTumorExpression) ;
plot(brcaExampleTumorExpression)
```

---

condorcetRanking *condorcetRanking*

Description

condorcetRanking determines the aggregate rank of a gene over a population.The condorcet rank is used to determine the overall strength of driver gene compared to other drivers by having patients "vote" for a winner using pairwise comparisions.

Usage

```
condorcetRanking(scoreMatrix, mutationMatrix, pen = 0.85)
```

Arguments

scoreMatrix,   a matrix containing all the given DawnRank scores per patient. Rows are genes, columns are patients

mutationMatrix,

the mutation matrix. Rows are genes, columns are patients

pen,   the penality parameter in the condorcet algorithm for missing data. Default 0.85

Value

the ranks. A list of 2 including a [[1]] a matrix of all pairwise comparisons, [[2]] the final rankings based on the Condorcet score

Examples

```
###using a small subset of the TCGA dataset and a small KEGG
###gene interaction network,
###We will show how to get DawnRank Results

library(DawnRank)

#load the mutation data
data(brcaExampleMutation)

#load the tumor expression data
data(brcaExampleTumorExpression)

#load the normal expression data
data(brcaExampleNormalExpression)

#load the pathway data
data(brcaExamplePathway)

#load the gold standard
data(goldStandard)

#normalize the tumor and normal data to get the differential expression
normalizedDawn<-DawnNormalize(tumorMat=brcaExampleTumorExpression,
normalMat=brcaExampleNormalExpression)
```

```
#get the DawnRank Score
dawnRankScore<-DawnRank(adjMatrix=brcaExamplePathway,
mutationMatrix=brcaExampleMutation,expressionMatrix=normalizedDawn,
mu=3,goldStandard=goldStandard)

#look at the DawnRank scores for a few patients
dawnRankFrame<-dawnRankScore[[3]]
head(dawnRankFrame)

#get the aggregate DawnRank scores
aggregateDawnRankScore<-condorcetRanking(scoreMatrix=dawnRankScore[[2]],
mutationMatrix=brcaExampleMutation)

#look at top 10 ranked genes
top10<-aggregateDawnRankScore[[2]][1:10]
top10

#get the individual cutoff for patient TCGA-A2-A04P
dawnRankFrame$isCGC<-dawnRankFrame$isGoldStandard
library(maxstat)
significance<-patspeccutoff(patient="TCGA-A2-A04P",ms=dawnRankFrame,
default=95)
```

Dawn *Dawn*

Description

The shell of the DawnRank algorithm for single patient. Dawn determines rankings of mutated genes for only one patient. Dawn is called in a loop by DawnRank.

Usage

```
Dawn(adjMatrix, expressionVector, mutationVector,
  damping, maxit = 100, epsilon = 1e-04,
  goldStandard = NULL, patientTag = "defaultPatient")
```

Arguments

adjMatrix, the adjacency
expressionVector,
        the normalized expression vector
mutationVector,
        a logical vector containing mutation information
damping, the damping vector
maxit, the maximum number of iterations to use, default 100
epsilon, the lower magnitude cutoff, default 0.0001
goldStandard, A list of common driver genes, used as a comparison. This is optional, default=NULL
patientTag, an index for the patients

Value

the ranks. A list of 3 including a [[1]] output of all the ranks, [[2]] mutated ranks, [[3]] the steps of convergence

---

dawnDamping   *dawnDamping*

Description

dawnDamping calculates the damping factor used in DawnRank using the DirichletRank dynamic damping factor.DawnDamping can be called by DawnRank should the user supply the mu parameter.

Usage

```
dawnDamping(adjMatrix, mu)
```

Arguments

adjMatrix,   the adjacency matrix

mu,   the free parameter to fit

Value

the vector for individualized damping factors

DawnMatrix   *DawnMatrix*

Description

DawnMatrix creates a transition matrix for DawnRank where the transition probability of each node is determined by the number of incoming edges to that node.

Usage

```
DawnMatrix(adjMatrix)
```

Arguments

adjMatrix,   the adjacency matrix

Value

the transition matrix

References

Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry (1999) The PageRank Citation Ranking: Bringing Order to the Web. Technical Report.

```
DawnNormalize          DawnNormalize
```

Description

DawnNormalize takes a tumor and normal expression matrix and returns a standardized absolute differential expression matrix. The resulting data is mapped to a standard normal distribution.

Usage

```
DawnNormalize(tumorMat, normalMat)
```

Arguments

tumorMat,   a matrix representing the tumor expression

normalMat,   a matrix representing the normal expression

Value

the differential expression matrix, it is absolute standarized matrix

Examples

```
###using a small subset of the TCGA dataset and a small KEGG
###gene interaction network,
###We will show how to get DawnRank Results

library(DawnRank)

#load the mutation data
data(brcaExampleMutation)

#load the tumor expression data
data(brcaExampleTumorExpression)

#load the normal expression data
data(brcaExampleNormalExpression)

#load the pathway data
data(brcaExamplePathway)

#load the gold standard
data(goldStandard)

#normalize the tumor and normal data to get the differential expression
normalizedDawn<-DawnNormalize(tumorMat=brcaExampleTumorExpression,
normalMat=brcaExampleNormalExpression)

#get the DawnRank Score
dawnRankScore<-DawnRank(adjMatrix=brcaExamplePathway,
mutationMatrix=brcaExampleMutation,expressionMatrix=normalizedDawn,
mu=3,goldStandard=goldStandard)

#look at the DawnRank scores for a few patients
```

```
dawnRankFrame<-dawnRankScore[[3]]
head(dawnRankFrame)

#get the aggregate DawnRank scores
aggregateDawnRankScore<-condorcetRanking(scoreMatrix=dawnRankScore[[2]],
mutationMatrix=brcaExampleMutation)

#look at top 10 ranked genes
top10<-aggregateDawnRankScore[[2]][1:10]
top10

#get the individual cutoff for patient TCGA-A2-A04P
dawnRankFrame$isCGC<-dawnRankFrame$isGoldStandard
library(maxstat)
significance<-patspeccutoff(patient="TCGA-A2-A04P",ms=dawnRankFrame,
default=95)
```

DawnRank   *DawnRank*

Description

This is the main method of the DawnRank method. DawnRank calculates the importance of a gene in a pathway through its connectivity to downstream genes and the differential expression of the downstream genes.

Usage

```
DawnRank(adjMatrix, expressionMatrix, mutationMatrix,
  mu = 20, maxit = 100, epsilon = 1e-04,
  goldStandard = NULL)
```

Arguments

adjMatrix,   the adjacency matrix

expressionMatrix,
                  the normalized expression matrix (multiple patients)

mutationMatrix,
                  a logical matrix containing mutation information

mu,   the proposed free parameter

maxit,   the maximum number of iterations to use, default 100

epsilon,   the lower magnitude cutoff, default 0.0001

goldStandard,   A list of common driver genes, used as a comparison. This is optional, default=NULL

Value

the ranks. A list of 3 including a [[1]] output of all the ranks, [[2]] output of all the ranks (percentile), [[3]] mutated ranks, [[4]] the steps of convergence

Examples

```
###using a small subset of the TCGA dataset and a small KEGG
###gene interaction network,
###We will show how to get DawnRank Results

library(DawnRank)

#load the mutation data
data(brcaExampleMutation)

#load the tumor expression data
data(brcaExampleTumorExpression)

#load the normal expression data
data(brcaExampleNormalExpression)

#load the pathway data
data(brcaExamplePathway)

#load the gold standard
data(goldStandard)

#normalize the tumor and normal data to get the differential expression
normalizedDawn<-DawnNormalize(tumorMat=brcaExampleTumorExpression,
normalMat=brcaExampleNormalExpression)

#get the DawnRank Score
dawnRankScore<-DawnRank(adjMatrix=brcaExamplePathway,
mutationMatrix=brcaExampleMutation,expressionMatrix=normalizedDawn,
mu=3,goldStandard=goldStandard)

#look at the DawnRank scores for a few patients
dawnRankFrame<-dawnRankScore[[3]]
head(dawnRankFrame)

#get the aggregate DawnRank scores
aggregateDawnRankScore<-condorcetRanking(scoreMatrix=dawnRankScore[[2]],
mutationMatrix=brcaExampleMutation)

#look at top 10 ranked genes
top10<-aggregateDawnRankScore[[2]][1:10]
top10

#get the individual cutoff for patient TCGA-A2-A04P
dawnRankFrame$isCGC<-dawnRankFrame$isGoldStandard
library(maxstat)
significance<-patspeccutoff(patient="TCGA-A2-A04P",ms=dawnRankFrame,
default=95)
```

---

goldStandard   *goldStandard*

Description

The Cancer Gene Census (CGC) Gold standard for common driver genes. The CGC is maintained by COSMIC.

Usage

```
data(goldStandard)
```

Format

The format is: chr [1:147] "ABL1" "ABL2" "ACSL3" "AKAP9" "AKT1" "AKT2" "APC" "ARID1A" ...

Details

The Cancer Gene Census Gold standard for common driver genes.This is a vector

Source

http://cancer.sanger.ac.uk/cancergenome/projects/census/

References

Futreal P, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton M: A census of human cancer genes. Nature Reviews Cancer 2004, 4:177-183

Examples

```
data(goldStandard)
## maybe str(goldStandard) ; plot(goldStandard) ...
```

---

patspeccutoff   *patspeccuttoff*

Description

patspeccuttoff determines the score cutoffs to distinguish drivers and passenger mutations of a single patient using Maximially-Selected Rank Statistics. These cutoffs determine a threshold for each patient that maximizes the number of known genes that score above that threshold.

Usage

```
patspeccutoff(patient, ms, default = 0.95)
```

Arguments

patient,	the patient of interest
ms,	a dataframe containing dawn rank output (specifically, the third element of Dawn-Rank output)
default,	the default cutoff in case there are not enough common drivers to make a clear conclusion

Value

a data frame containing information regarding the genetic information of a specific patient, and the cutoff itself

References

Torsten Hothorn and Berthold Lausen. On maximally selected rank statistics. R News, 2(1):3<e2><80><93>5, 2002.

Examples

```
###using a small subset of the TCGA dataset and a small KEGG
###gene interaction network,
###We will show how to get DawnRank Results

library(DawnRank)

#load the mutation data
data(brcaExampleMutation)

#load the tumor expression data
data(brcaExampleTumorExpression)

#load the normal expression data
data(brcaExampleNormalExpression)

#load the pathway data
data(brcaExamplePathway)

#load the gold standard
data(goldStandard)

#normalize the tumor and normal data to get the differential expression
normalizedDawn<-DawnNormalize(tumorMat=brcaExampleTumorExpression,
normalMat=brcaExampleNormalExpression)

#get the DawnRank Score
dawnRankScore<-DawnRank(adjMatrix=brcaExamplePathway,
mutationMatrix=brcaExampleMutation,expressionMatrix=normalizedDawn,
mu=3,goldStandard=goldStandard)

#look at the DawnRank scores for a few patients
dawnRankFrame<-dawnRankScore[[3]]
head(dawnRankFrame)

#get the aggregate DawnRank scores
aggregateDawnRankScore<-condorcetRanking(scoreMatrix=dawnRankScore[[2]],
mutationMatrix=brcaExampleMutation)

#look at top 10 ranked genes
top10<-aggregateDawnRankScore[[2]][1:10]
top10

#get the individual cutoff for patient TCGA-A2-A04P
dawnRankFrame$isCGC<-dawnRankFrame$isGoldStandard
library(maxstat)
```

```
significance<-patspeccutoff(patient="TCGA-A2-A04P",ms=dawnRankFrame,
default=95)
```