Package ‘csSAM’

February 19, 2015

Type Package
Title csSAM - cell-specific Significance Analysis of Microarrays
Version 1.2.4
Date 2011-10-08
Author Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang
Maintainer Shai Shen-Orr <shenorr@technion.ac.il>
Description Cell-type specific differential expression of a microarray experiment of heterogeneous tissue samples, using SAM.
License LGPL
LazyLoad yes
Depends R (>= 2.15), compiler
Collate 'csfit.R' 'csSAM-package.R' 'csSAM.R' 'csSamWrapper.R'
    'fdrCsSAM.R' 'fdrSAM.R' 'findSigGene.R' 'make.monotone.R'
    'plotCsSAM.R' 'runSAM.R' 'ttest.func.R' 'varr.R'
NeedsCompilation no
Repository CRAN
Date/Publication 2013-05-13 17:39:59

R topics documented:

  csSAM-package .................................................. 2
  csfit .......................................................... 4
  csSAM .......................................................... 5
  csSamWrapper ................................................... 6
  fdrCsSAM ......................................................... 7
  fdrSAM .......................................................... 9
  findSigGene ..................................................... 10
  plotCsSAM ....................................................... 11
  runSAM .......................................................... 11

Index 13
csSAM-package

Cell-specific Differential Expression (csSAM)

Description

SAM for Cell-specific Differential Expression SAM.

Details

Package:
Type:
Version: 1.2
Date: 2011-10-08
License: LGPL
LazyLoad: yes

Tissues are often made up of multiple cell-types. Each with its own functional attributes and molecular signature. Yet, the property individual in the group, it is the amount contributed by that cell type to the overall measured expression on the array.

Key functions for this package:
- csSamWrapper: Single wrapper function performs all functionality.
- csfit: For deconvolving the average cell-type specific expression.
- csSAM: For calculating the contrast between every pair of cells being compared between the two groups.
- fdrCsSAM: Estimate the false discovery rate for each cell-type specific comparison.
- findSigGenes: Identifies the list of differentially expressed genes in a given cell-type at a given FDR cutoff.
- plotCsSAM: Plots a fdr plot of the results.

Additional functions exists (runSAM and fdrSAM to contrast csSAM with the tissue heterogeneity ignorant SAM).

Author(s)

Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang
Maintainer: Shai Shen-Orr <shenorr@stanford.edu>

References


Examples

```r
library("csSAM")
##
## Generate random dataset
##
set.seed(143)
```
k <- 5 # number of cell types
gg <- 500 # number of genes
pp <- 20 # number of samples
nndiff <- 100 # number of genes differentially expressed

# true cell-specific signatures
HH1 <- matrix(rnorm(5*gg), ncol=gg)
HH2 <- HH1
# create differential expression for 3rd cell type

# cell frequency matrix per sample
cc <- matrix(runif(pp*k), ncol=k)
cc <- t(scale(t(cc), center=FALSE, scale=rowSums(cc)))
colnames(cc) <- paste('cellType', 1:ncol(cc), sep='')

# global expression matrix
GG <- rbind(cc[1:10],] %*% HH1, cc[11:p,] %*% HH2) + matrix(rnorm(p*gg), ncol=gg)
# sample classes (2 groups)
y <- gl(2, p/2)

fileName = "Example File.pdf"

# Now run, either using the wrapper
# NB: more permutations would be needed for real data
deconvResults = csSamWrapper(G, cc, y, nperms = 50, alternative = "two.sided"
, standardize = TRUE
, medianCenter = TRUE
, fileName = fileName)

# Or by calling each function independently:
# this is useful if you want to perform only cell-specific expression
# without differential expression.
## Not run:
numset = nlevels(y)
n <- summary(y, maxsum=Inf) # number of samples in each class
numgene = ncol(G)
umcell = ncol(cc)
geneID = colnames(G)
cellID = colnames(cc)

deconv <- list()
# run analysis
for (curset in levels(y))
deconv[[curset]] = csfit(cc[y==curset,], G[y==curset])

rhat <- array(dim = c(numcell, numgene))
rhat[, ] <- csSAM(deconv[[1]]$ghat, deconv[[1]]$se,
n[1], deconv[[2]]$ghat, deconv[[2]]$se, n[2],
standardize=TRUE, medianCenter=TRUE, nonNeg=TRUE)
tt.sam <- runSAM(G, y)
falseDiscovR <- fdrCsSAM(G, cc, y, n, numcell, numgene, rhat, nperms = 200, standardize=TRUE, alternative='two.sided', medianCenter=TRUE, nonNeg=TRUE)
falseDiscovRSAM <- fdrSAM(G, y, nperms=200, alternative = 'two.sided', tt.sam)
sigGene <- findSigGene(G, cc, y, rhat, falseDiscovR)

plotCsSAM(falseDiscovR, falseDiscovRSAM, alternative='two.sided', cellID, numcell, fileName)
print (falseDiscovR$fdr.g[, ])

## End(Not run)

---

csfit

**csfit**: Deconvolution from Known Cell Proportions

**Description**

Deconvolves cell-specific expression using least-squares fit. Input is the heterogeneous sample gene expression of a group of samples and the matching cell-frequencies of the sample. The lower limit for the number of samples needed to deconvolving the cell-specific expression of N cell-types is N+1. For a single color array - the result could be interpreted as the average expression level of a given gene in a cell-type of that group. Multiplied by the frequency of a given cell-type in an individual in the group, it is the amount contributed by that cell type to the overall measured expression on the array.

**Usage**

csfit(cc, G, logRm = FALSE, logBase = 2)

**Arguments**

- **G**: Matrix of gene expression, columns ordered in the same order at the cell-frequency matrix (n by g, n samples, g genes)
- **cc**: Matrix of cell-frequency. (n by k, n samples, k cell-types)
- **logRm**: Exponentiate data for deconvolution stage. Default is FALSE
- **logBase**: Base of logarithm used to determine exponentiation factor. Default is 2

**Value**

A list with three attributes:

- **ghat**: A matrix of cell-specific expression for each gene as derived from the coefficients of the fit. (Size: k by g, k cell types, gp genes)
- **se**: Standard error of the fit coefficients
- **residuals**: The individual sample residuals.

**Author(s)**

Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang
References

Description
Computes the contrast between groups for the deconvolved cell-specific expression for each cell-type.

Usage
```
csSAM(ghat1, se1, n1, ghat2, se2, n2, standardize, mediancenter = TRUE, nonNeg = FALSE)
```

Arguments
- **ghat1**: Expression matrix of deconvolved cell-specific gene expression estimates for group 1.
- **se1**: Standard error group 1
- **n1**: Group 1 size
- **ghat2**: Expression matrix of deconvolved cell-specific gene expression estimates for group 2.
- **se2**: Standard error group 2
- **n2**: Group 2 size
- **standardize**: Standardize contrast values
- **medianCenter**: Median center rhat distributions for each cell-type
- **nonNeg**: Negative values not allowed such as in a single channel microarray. Zero them if negative (a conservative option)

Value
A matrix object with the result of contrasting the average cell-specific expression profile of the two groups, per cell-type (Size k by g where k is the number of cells and g is the number of genes).

Author(s)
Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang
References


---

**csSamWrapper**

**csSamWrapper function - performs entire functionality**

**Description**

csSamWrapper function - performs entire functionality

**Usage**

```r
csSamWrapper(G, cc, y, nperms = 200,
  alternative = "two.sided", standardize = TRUE,
  medianCenter = TRUE, logRm = FALSE, logBase = 2,
  nonNeg = TRUE, fileName = "csSAMout.pdf")
```

**Arguments**

- **G**
  - Matrix of gene expression, columns ordered in the same order at the cell-frequency matrix (n by g, n samples, g genes)
- **cc**
  - Matrix of cell-frequency. (n by k, n samples, k cell-types)
- **y**
  - A numeric vector of group association of each sample. Either 1 or 2.
- **nperms**
  - The number of permutations to perform.
- **alternative**
  - `two.sided` less `greater`
- **standardize**
  - Standardize sample or not. Default is `TRUE`.
- **medianCenter**
  - Median center `rhat` distributions. Default is `TRUE`.
- **logRm**
  - Exponentiate data for deconvolution stage. Default is `FALSE`
- **logBase**
  - Base of logarithm used to determine exponentiation factor. Default is 2
- **nonNeg**
  - For single channel arrays. Set any cell-specific expression estimated as negative, to a ceiling of 0. It is conservative in its study of differential expression. Default is `FALSE`.
- **fileName**
  - PDF file containing plots of FDR vs. number of genes called for whole tissue comparison (via SAM) as well as each cell-type (by csSAM)
fdrCsSAM

Value

Returns a list containing:

deconv A list object containing a fit (cell-type specific expression) for each group. Each element in the list is an object returned by csFit.

fdr.csSAM A list output of the fdrCsSAM function.

fdr.SAM A list output of the fdrSAM function.

sigGene.csSAM A list of significant genes.

fileName The filename into which the FDR plots are dumped.

Author(s)

Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang

References


See Also

csfit.csSAM,fdrCsSAM,plotCsSAM

fdrCsSAM fdrCsSAM

Description

Estimates the false discovery rate for the identified cell-specific differences in gene expression.

Usage

fdrCsSAM(G, cc, y, n, numcell, numgene, rhat, nperms, alternative = "two.sided", standardize = TRUE, medianCenter = TRUE, logRm = FALSE, logBase = 2, nonNeg = FALSE)

Arguments

G Matrix of gene expression, columns ordered in the same order at the cell-frequency matrix (n by p, n samples, p genes)

cc Matrix of cell-frequency. (n by k, n samples, k cell-types)

y A numeric vector of group association of each sample. Either 1 or 2.
n  A numeric vector describing the number of samples in a group
numcell  The number of cell-types to consider
numgene  The number of genes being considered
rhat  The contrast in cell-type expression for each cell-type as observed between the two groups being compared.
nperms  The number of permutations to perform.
alternative  Type of test to conduct - choose between 'two.sided', 'greater', or 'less'
standardize  Standardize sample or not. Default is TRUE
medianCenter  Median center rhat distributions. Default is TRUE.
logRm  Exponentiate data for deconvolution stage. Default is FALSE
logBase  Base of logarithm used to determine exponentiation factor. Default is 2
nonNeg  For single channel arrays. Set any cell-specific expression estimated as negative, to a ceiling of 0. It is conservative in its study of differential expression. Default is FALSE.

Value
A list.

fdr.g  A matrix false discovery rates for csSAM comparison for each cell-type at different thresholds. A set of 100 thresholds is determined automatically from the data (k by 100, where k is number of cells).
avrhatperm  A matrix sized pXkXg which stores the contrast of a given gene g in cell type k in permutation p of the data.
rhatperm  A matrix sized pXkXg which stores the contrast of a given gene g in cell type k in permutation p of the data.
cutp.g  A matrix k by 100, where k is the number of cell types. Lists the 100 cutoff thresholds for each cell-type as determined automatically from the computed contrast.
rhat  A matrix object with the result of contrasting the average cell-specific expression profile of the two groups, per cell-type (Size k by g where k is the number of cells and g is the number of genes).
ncall.g  Number of genes called significant at the given cutoff threshold with a FDR matching that indicated in fdr.g

Author(s)
Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang

References
Description

Calculate the false discovery rate (FDR) by permutation for the group differences as calculated by SAM.

Usage

```
fdrSAM(G, y, nperms, tt.sam, alternative = "two.sided")
```

Arguments

- **G**: Matrix of gene expression, columns ordered in the same order as the cell-frequency matrix (n by g, n samples, g genes)
- **y**: A numeric vector of group association of each sample. Either 1 or 2.
- **nperms**: Number of permutations to run. User responsibility to the number appropriately fitting the sample size.
- **tt.sam**: Real group comparison t-test statistic value
- **alternative**: Type of test. Choices are ‘two.sided’, ‘greater’ or ‘less’

Value

- **fdr.sam**: A vector false discovery rates for SAM comparison at different thresholds. A set of 100 thresholds is determined automatically from the data.
- **ncall.sam**: Number of genes called significant at the given cutoff threshold with a FDR matching that indicated in fdr.sam
- **ttstar.sam**: A matrix listing the t statistic for each gene in each permutation. (p by g, p permutations, g genes)
- **siggene.sam**: A vector of length equal to the number of genes being considered. For each gene the estimated FDR is listed.

Author(s)

Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang

References

Description

Find the false discovery rate for each gene in each cell-type.

Usage

```r
cfindSigGene(G, cc, y, rhat, csSAMData)
```

Arguments

- `G`: Gene expression matrix of heterogenous tissue measurements
- `cc`: Matrix of cell-frequency measures per person
- `y`: Numeric group association of each sample. Either 1 or 2.
- `rhat`: Matrix of cell-specific contrasts for each gene in each cell-type as computed for the original group classification.
- `csSAMData`: List object returned from fdrCsSAM.

Value

A matrix size k by g where k is the number of cell-types and g is the number of genes. For each cell in the matrix, listed is the FDR of the gene for a difference in a given cell-type.

Author(s)

Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang

References

plotCsSAM

---

**Description**

Plots the # of genes called significant at a given false discovery rate for the SAM (heterogenous tissue) comparison, and for each of the contrasted cell-types using csSAM

**Usage**

```r
plotCsSAM(csSAMdata, SAMdata, alternative, cellID, numcell, fileName)
```

**Arguments**

- `csSAMdata`: List object output of the fdrCsSAM function
- `SAMdata`: List object output of the fdrSAM function
- `alternative`: Type of test conducted. Will appear in plot title.
- `cellID`: Label for each cell-type
- `numcell`: Number of different cell-types being considered.
- `fileName`: Name of output pdf file.

**Author(s)**

Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang

**References**


---

runSAM

---

**Description**

A lightweight version of the SAM algorithm, only performs two group comparison with equal deltas on each tail

**Usage**

```r
runSAM(G, y, s0.sam = NULL, stand.r = TRUE)
```
Arguments

G  Matrix of gene expression, columns ordered in the same order at the cell-frequency matrix (n by p, n samples, p genes)

y  Numeric group association of each sample. Either 1 or 2.

s0.sam  Input or computed value of SAM exchangeability factor. Default is determined automatically

standNr  Median center and standardize arrays. Default is TRUE.

Author(s)

Shai Shen-Orr, Rob Tibshirani, Narasimhan Balasubramanian, David Wang

References

Index

csfit, 4, 7
csSAM, 5, 7
csSAM-package, 2
cssamWrapper, 6

fdrCsSAM, 7, 7
fdrSAM, 9
findSigGene, 10

plotCsSAM, 7, 11
runSAM, 11