

Package ‘DisHet’

May 25, 2018

Type Package

Title Estimate the Gene Expression Levels and Component Proportions of the Normal, Stroma (Immune) and Tumor Components of Bulk Tumor Samples

Version 1.0.0

Date 2018-05-25

Author Tao Wang

Maintainer Tao Wang <tao.wang@utsouthwestern.edu>

Depends R (>= 2.15.0), gtools, matrixStats

Description Model cell type heterogeneity of bulk renal cell carcinoma. The observed gene expression in bulk tumor sample is modeled by a log-normal distribution with the location parameter structured as a linear combination of the component-specific gene expressions.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2018-05-25 21:40:01 UTC

R topics documented:

DisHet	2
eTME_signatures	3
StromaExp	5
Index	6

Description

This function performs dissection of bulk sample gene expression using matched normal and tumorgraft RNA-seq data. It outputs the final proportion estimates of the three components for all patients.

The patient-specific dissection proportion estimates are saved in a 3-by-k matrix named "rho", where k is the number of patients. The 3 rows of "rho" matrix correspond to the tumor, normal, stroma components in order. That is, the proportion estimate of tumor component for patient i is stored in rho[1,i]; the normal component proportion estimate of this patient is stored in rho[2,i]; and stroma component proportion in rho[3,i].

Usage

```
DisHet(exp_T,exp_N,exp_G, save=TRUE, MCMC_folder,
       n_cycle=10000, save_last=500, mean_last=200, dirichlet_c=1, S_c=1, rho_small=1e-2,
       initial_rho_S=0.02, initial_rho_G=0.96, initial_rho_N=0.02)
```

Arguments

exp_T	Gene expression in bulk RNA-seq samples. The rows correspond to different genes. The columns correspond to different patients.
exp_N	Gene expression in the corresponding normal samples. The rows list the same set of genes as in exp_G. The columns correspond to patients matched with exp_T.
exp_G	Gene expression in the corresponding tumor samples. The rows list the same set of genes as in exp_G. The columns correspond to patients matched with exp_T.
save	When save==TRUE, as in default, all component proportion estimates during MCMC iterations can be saved into a user-specified directory using the "MCMC_folder" argument.
MCMC_folder	Directory for saving the estimated mixture proportion matrix updates during MCMC iterations. The default setting is to create a "DisHet" folder under the current working directory.
n_cycle	Number of MCMC iterations(chain length). The default value is 10,000.
save_last	Save the rho matrix updates for the last "save_last" Number of MCMC iterations. The default value is 500.
mean_last	Calculate the final proportion estimate matrix using the last "mean_last" number of MCMC iterations. The default value is 200.
dirichlet_c	Stride scale in sampling rho. Larger value leads to smaller steps in sampling rho. The default value is 1.
S_c	Stride scale in sampling Sij. Larger value leads to larger steps in sampling Sij. The default value is 1.

rho_small	The smallest rho updates allowed during MCMC. The default is 1e-2. This threshold is set to help improve numerical stability of the algorithm.
initial_rho_S	Initial value of the proportion estimate for the stroma component. The default value is 0.02.
initial_rho_G	Initial value of the proportion estimate for the tumor component. The default value is 0.96.
initial_rho_N	Initial value of the proportion estimate for the normal component. The default value is 0.02.

Details

Un-logged expression values should be used in exp_N/T/G matrices, and their rows and columns must match each other corresponding to the same set of genes and patients.

The values specified for "initial_rho_S", "initial_rho_G", and "initial_rho_S" all have to be positive. If the three proportion initials are not summing to 1, normalization is performed automatically to force the sum to be 1.

Examples

```
load(system.file("example/example_data.RData", package="DisHet"))
exp_T <- exp_T[1:200,]
exp_N <- exp_N[1:200,]
exp_G <- exp_G[1:200,]

rho <- DisHet(exp_T, exp_N, exp_G, save=FALSE, n_cycle=200, mean_last=50)
```

eTME_signatures *Empirically-defined Immune Signature Genes in RCC Bulk Tumor*

Description

Based on DisHet analysis of 35 RCC trio RNA-Seq data, we defined immune-specific genes with empirical evidence, named eTME, for empirically-defined immune signature of tumor. Using eTME, we refined previously published Immunome signatures. We also assigned other eTME genes to specific immune cell types using the BLUEPRINT project data. These two sets of refined gene signatures were consolidated and documented in the DisHet R package as the "eTME" immune cell gene signatures.

Usage

```
data("eTME_signatures")
```

Format

A List contains 2 lists of signatures:the 1st list "signatures_gene" contains the signatures on gene level, and the 2nd list "signatures_mRNA" contains the signatures on mRNA level. Both lists contain 25 items/vectors.

'**M2 macrophages**' : a vector of genes/mRNA that are abundantly expressed in M2 macrophages

'**M1 macrophages**' : a vector of genes/mRNA that are abundantly expressed in M1 macrophages

Macrophages : a vector of genes/mRNA that are abundantly expressed in Macrophages

Monocytes : a vector of genes/mRNA that are abundantly expressed in Monocytes

'**B cells**' : a vector of genes/mRNA that are abundantly expressed in B cells

'**CD8 T cells**' : a vector of genes/mRNA that are abundantly expressed in CD8 T cells

'**T cells**' : a vector of genes/mRNA that are abundantly expressed in T cells

'**Dendritic cells**' : a vector of genes/mRNA that are abundantly expressed in Dendritic cells

'**CD56dim NK cells**' : a vector of genes/mRNA that are abundantly expressed in CD56dim NK cells

'**CD56bright NK cells**' : a vector of genes/mRNA that are abundantly expressed in CD56bright NK cells

'**NK cells**' : a vector of genes/mRNA that are abundantly expressed in NK cells

'**Endothelial cells**' : a vector of genes/mRNA that are abundantly expressed in Endothelial cells

Eosinophils : a vector of genes/mRNA that are abundantly expressed in Eosinophils

Neutrophils : a vector of genes/mRNA that are abundantly expressed in Neutrophils

'**Treg cells**' : a vector of genes/mRNA that are abundantly expressed in Treg cells

'**Th1 cells**' : a vector of genes/mRNA that are abundantly expressed in Th1 cells

'**Th2 cells**' : a vector of genes/mRNA that are abundantly expressed in Th2 cells

'**Tfh cells**' : a vector of genes/mRNA that are abundantly expressed in Tfh cells

'**Th cells**' : a vector of genes/mRNA that are abundantly expressed in Th cells

aDCs : a vector of genes/mRNA that are abundantly expressed in aDCs

iDCs : a vector of genes/mRNA that are abundantly expressed in iDCs

pDCs : a vector of genes/mRNA that are abundantly expressed in pDCs

'**Mast cells**' : a vector of genes/mRNA that are abundantly expressed in Mast cells

'**Tm cells**' : a vector of genes/mRNA that are abundantly expressed in Tm cells

Pericytes : a vector of genes/mRNA that are abundantly expressed in Pericytes

Examples

```
data(eTME_signatures)
eTME_signatures$signatures_gene$Macrophages
eTME_signatures$signatures_mRNA$Macrophages
eTME_signatures$signatures_gene`T cells`
eTME_signatures$signatures_mRNA`T cells`
```

StromaExp

Stroma (Immune) Component Gene Expression Estimation

Description

This function estimate the stroma component gene expression profiles of all patients, using the proportion estimates obtained from function `DisHet`. The estimates are stored in a p-by-k matrix, where p is the number of genes and k is the number of patients. The order of genes and the order of patients are the same as in the input bulk sample expression matrix.

Usage

```
StromaExp(exp_T, exp_N, exp_G, rho)
```

Arguments

<code>exp_T</code>	Gene expression in bulk RNA-seq samples. The rows correspond to different genes. The columns correspond to different patients.
<code>exp_N</code>	Gene expression in the corresponding normal samples. The rows list the same set of genes as in <code>exp_G</code> . The columns correspond to patients matched with <code>exp_T</code> .
<code>exp_G</code>	Gene expression in the corresponding tumor samples. The rows list the same set of genes as in <code>exp_G</code> . The columns correspond to patients matched with <code>exp_T</code> .
<code>rho</code>	Output from function <code>DisHet</code> : the patient-specific proportion estimates corresponding to tumor, normal, stroma components in order.

Examples

```
load(system.file("example/example_data.RData", package="DisHet"))
exp_T <- exp_T[1:200,]
exp_N <- exp_N[1:200,]
exp_G <- exp_G[1:200,]

rho <- DisHet(exp_T, exp_N, exp_G, save=FALSE, n_cycle=200, mean_last=50)
S <- StromaExp(exp_T, exp_N, exp_G, rho)
```

Index

*Topic **datasets**

eTME_signatures, [3](#)

DisHet, [2](#)

eTME_signatures, [3](#)

StromaExp, [5](#)