Package 'InteRD'

October 12, 2022

Type Package

Title The Integrated and Robust Deconvolution

Version 0.1.1

Description We developed the Integrated and Robust Deconvolution algorithm to infer cell-type proportions from target bulk RNA-seq data. This package is able to effectively integrate deconvolution results from multiple scRNA-seq datasets and calibrates estimates from reference-based deconvolution by taking into account extra biological information as priors. Moreover, the proposed algorithm is robust to inaccurate external information imposed in the deconvolution system.

License Artistic-2.0

Encoding UTF-8

RoxygenNote 7.2.1

URL https://github.com/chencxxy28/InteRD

BugReports https://github.com/chencxxy28/InteRD/issues

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

biocViews

Imports Rcpp (>= 0.11.0), limSolve, cowplot, ggplot2, pheatmap, stats,DescTools, mgcv, reshape2

Depends R (>= 3.5.0), Biobase

Config/testthat/edition 3

NeedsCompilation no

Author Chixiang Chen [cre, aut],
Yuk Yee Leung [aut],
Matei Lonita [aut],
Li-San Wang [aut],
Mingyao Li [aut]

Maintainer Chixiang Chen <chencxxy@hotmail.com>

Repository CRAN

Date/Publication 2022-08-12 07:20:11 UTC

R topics documented:

`evaluate` *Evaluation for estimated cell type proportions*

Description

Several evaluation metrics are provided, such as mean absolute deviance ('MAD'), Kendall-tau correlation coefficient ('Ken'), Pearson correlation coefficient ('Cor'), given true cell type proportions.

Usage

```
evaluate(est.prop,true.prop)
```

Arguments

`est.prop`	The estimated cell type proportions.
`true.prop`	The True cell type proportions

Value

Cell-type level evaluations based on MAD, Ken, and Pearson ('cell.type.eva'), and overall evaluations based on averaged MAD, Ken, and Pearson ('all.eva').

Examples

```
##read data
library(InteRD)
readRDSFromWeb<-function(ref) {readRDS(gzcon(url(ref)))}
urlremote<-"https://github.com/chencxxy28/Data/raw/main/data_InteRD/"
pseudo.seger<-readRDSFromWeb(paste0(urlremote,"pseudo.seger.rds"))
true_p<-readRDSFromWeb(paste0(urlremote,"true_p.rds"))
SCDC_ENSEMBLE_MAD<-readRDSFromWeb(paste0(urlremote,"SCDC_ENSEMBLE_MAD_seger.rds"))
evaluate(SCDC_ENSEMBLE_MAD,true_p)$all.eva
```

generateBulk *Pseudo bulk data generation function*

Description

This function generates a pseudo bulk samples by random sampled number of cells per subject.

Usage

```
generateBulk(eset,ct.varname,sample,disease = NULL,ct.sub,prop_mat = NULL,
nbulk = 50,samplewithRep = FALSE,low_s = 0.3,upp_s = 0.7)
```

Arguments

eset	The 'ExpressionSet' object for single cells.
ct.varname	Variable name for 'cell types'.
sample	Variable name for subject/samples.
disease	Indicate the health condition of subjects.
ct.sub	A subset of cell types that are selected to construct pseudo bulk samples. If NULL, then all cell types are used.
prop_mat	Manually input the cell-type proportion for pseudo bulk samples.
nbulk	The number of pseudo bulk samples to be constructed.
samplewithRep	Logical, randomly sample single cells with replacement. Default is F.
low_s	Lower support a for uniform distribution U[a,b].
upp_s	Upper support b for uniform distribution U[a,b].

Value

Pseudo bulk samples in the format of 'ExpressionSet', and the true cell-type proportions.

Examples

```
##read data
library(InteRD)
readRDSFromWeb<-function(ref) {readRDS(gzcon(url(ref)))}
urlremote<-"https://github.com/chencxxy28/Data/raw/main/data_InteRD/"
seger<-readRDSFromWeb(paste0(urlremote,"segerstolpe.rds"))

##generate a pseudo bulk data with two samples
set.seed(1234567)
pseudo.seger<-generateBulk(seger[["sc.eset.qc"]], ct.varname = "cluster",
sample = "sample", ct.sub = c("alpha","beta","delta","gamma"),
nbulk = 2, low_s = 0.3, upp_s = 0.7)
```

InteRD.predict.prop *Extract the estimated proportions from InteRD*

Description

This function extract estimated cell type proportions via InteRD1 and InteRD2.

Usage

```
InteRD.predict.prop(InteRD.output)
```

Arguments

InteRD.output	An object from InteRD1 or InteRD2.

Value

Estimated cell type proportions from InteRD.

Examples

```
##read data
library(InteRD)
readRDSFromWeb<-function(ref) {readRDS(gzcon(url(ref)))}
urlremote<-"https://github.com/chencxxy28/Data/raw/main/data_InteRD/"
InteRD1.output<-readRDSFromWeb(paste0(urlremote,"InteRD1.output.rds"))
lambda_option<-0
cell_type_unique<-c("alpha","beta","delta","gamma")
InteRD1<-InteRD.predict.prop(InteRD.output=InteRD1.output)
```

InteRD1 *The InteRD1 estimate from reference ensemble*

Description

This function provides a reference-based deconvolution by resembling all estimated cell-type proportions based on each reference set.

Usage

```
InteRD1(bulk.data,list_marker,cell_type_unique,comb_used,
lambda_option,tol=1e-06)
```

Arguments

bulk.data	The 'ExpressionSet' object for a target bulk data.
list_marker	A list of pre-specified marker genes corresponding to each cell type.
cell_type_unique	A list of cell types. It should match the order in list.marker.
comb_used	A list of pre-estimated cell type proportions based on different references.
lambda_option	A sequence of values for the tuning parameter.
tol	A tolerance value for convergence. The default is 1e-06

Value

A list containing estimated cell type proportions corresponding to each tuning value, named 'est'; and a sequence of goodness-of-fit values corresponding to each tuning value, named 'metrics'. The smaller the better; and a list of weights corresponding to each tuning value, named 'weights_list'.

Examples

```
##read data
library(InteRD)
readRDSFromWeb<-function(ref) {readRDS(gzcon(url(ref)))}
urlremote<-"https://github.com/chencxxy28/Data/raw/main/data_InteRD/"
pseudo.seger<-readRDSFromWeb(paste0(urlremote,"pseudo.seger.rds"))
comb<-readRDSFromWeb(paste0(urlremote,"comb_seger.rds"))
list_marker<-readRDSFromWeb(paste0(urlremote,"list_markerbaron20.rds"))
lambda_option<-0
cell_type_unique<-c("alpha","beta","delta","gamma")
InteRD1.output<-InteRD1(bulk.data =pseudo.seger,list_marker,
cell_type_unique,comb_used=comb,lambda_option,tol=1e-02)
InteRD1<-InteRD.predict.prop(InteRD.output=InteRD1.output)
```

InteRD2 *The InteRD2 estimate*

Description

This function provides a robust deconvolution framework to integrate information from scRNA-seq references, marker genes, and prior biological knowledge.

Usage

```
InteRD2(bulk.data,list_marker,cell_type_unique,comb_sampled,ave_est,ave_sd,
lambda_option,tol=0.0005)
```

Arguments

bulk.data	The 'ExpressionSet' object for a target bulk data.
list_marker	A list of pre-specified marker genes corresponding to each cell type.
cell_type_unique	A list of cell types. It should match the order in list.marker.
comb_sampled	A pre-specified cell type proportions for the target bulk data, which could be obtained from reference-based deconvolution approach.
ave_est	A pre-specified population-level cell type proportions, which could be obtained from single-cell RNA-seq and external expression data from different studies, species, or data types
ave_sd	A pre-specified standard deviation for cell-type proportion estimation. The default is 1 for each cell type.
lambda_option	A sequence of values for the tuning parameter.
tol	A tolerance value for convergence. The default is 0.0005.

Value

A list containing estimated cell type proportions corresponding to each tuning value, named 'est'; and a sequence of goodness-of-fit values corresponding to each tuning value, named 'metrics'. The smaller the better.

Examples

```
##read data
library(InteRD)
readRDSFromWeb<-function(ref) {readRDS(gzcon(url(ref)))}
urlremote<-"https://github.com/chencxxy28/Data/raw/main/data_InteRD/"
pseudo.seger<-readRDSFromWeb(paste0(urlremote,"pseudo.seger.rds"))
InteRD1<-readRDSFromWeb(paste0(urlremote,"InteRD1.rds"))
ave_est<-readRDSFromWeb(paste0(urlremote,"ave_est.rds"))
ave_sd<-readRDSFromWeb(paste0(urlremote,"ave_sd.rds"))
list_marker<-readRDSFromWeb(paste0(urlremote,"list_markerbaron20.rds"))
lambda_option<-0
cell_type_unique<-c("alpha","beta","delta","gamma")
lambda_option<-10e+05
InteRD2.output<-InteRD2(bulk.data=pseudo.seger,list_marker,cell_type_unique,
comb_sampled=InteRD1,ave_est,ave_sd,lambda_option=lambda_option,tol=0.01)
InteRD2<-InteRD.predict.prop(InteRD.output=InteRD2.output)
```

---

pop.ct.prop.scRNA *Calculate the population-level cell type proportions from a single-cell data.*

Description

Calculate population-level cell type proportions from single-cell data.

Usage

```
pop.ct.prop.scRNA(scRNA,cluster="cluster",sample="sample",cell_type_unique)
```

Arguments

scRNA	The 'ExpressionSet' object for single-cell data.
cluster	The character string specifying the variable name for cell types. The default is "cluster".
sample	The character string specifying the variable name for subject/samples. The default is "sample".
cell_type_unique	A vector of cell types. It should match the order in list.marker.

Value

The population-level cell type proportions ('pop.ct.prop') and corresponding standard deviations ('pop.ct.sd').

Examples

```
##read data
library(InteRD)
readRDSFromWeb<-function(ref) {readRDS(gzcon(url(ref)))}
urlremote<-"https://github.com/chencxxy28/Data/raw/main/data_InteRD/"
seger<-readRDSFromWeb(paste0(urlremote,"segerstolpe.rds"))
cell_type_unique<-c("alpha","beta","delta","gamma")
ave_est<-pop.ct.prop.scRNA(scRNA=seger[["sc.eset.qc"]],
cell_type_unique=cell_type_unique)$pop.ct.prop
ave_est
```

Ref_free *A reference-free deconvolution estimate*

Description

This function provides a reference-free deconvolution estimate, given a list of marker genes

Usage

```
Ref_free(bulk.data,list_marker,cell_type_unique,tol=0.001)
```

Arguments

bulk.data	The 'ExpressionSet' object for a target bulk data.
list_marker	A list of pre-specified marker genes corresponding to each cell type.
cell_type_unique	A list of cell types. It should match the order in 'list.marker'.
tol	A tolerance value for convergence. The default is 0.001.

Value

The estimated cell type proportions, named ‘est‘; and a goodness-of-fit value, named ‘metrics‘. The smaller the better.

Examples

```
##read data
library(InteRD)
readRDSFromWeb<-function(ref) {readRDS(gzcon(url(ref)))}
urlremote<-"https://github.com/chencxxy28/Data/raw/main/data_InteRD/"
pseudo.seger<-readRDSFromWeb(paste0(urlremote,"pseudo.seger.rds"))
list_marker<-readRDSFromWeb(paste0(urlremote,"list_markerbaron20.rds"))
cell_type_unique<-c("alpha","beta","delta","gamma")
ref_free.output<-Ref_free(bulk.data=pseudo.seger,list_marker=list_marker,
cell_type_unique=cell_type_unique,tol=0.01) #make tol=0.001
reffree<-InteRD.predict.prop(InteRD.output=ref_free.output)
```

Index