Package 'DeconRNASeq'

March 26, 2013

Type Package

Title Deconvolution of Heterogeneous Tissue Samples for mRNA-Seq data

Version 1.0.0

Date 2012-05-25

Author Ting Gong <tinggong@gmail.com>
Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Depends R (>= 2.14.0), limSolve, pcaMethods, ggplot2, grid

Maintainer Ting Gong <tinggong@gmail.com>

biocViews Bioinformatics, ExperimentData, RNAExpressionData

Description DeconSeq is an R package for deconvolution of heterogeneous tissues based on mRNA-Seq data. It modeled expression levels from heterogeneous cell populations in mRNA-Seq as the weighted average of expression from different constituting cell types and predicted cell type proportions of single expression profiles.

License GPL-2

R topics documented:

DeconRNASeq-package *package DeconRNASeq contains function "DeconRNASeq", implementing the decomposition of RNA-Seq expression profilings of heterogeneous tissues into cell/tissue type specific expression and cell type concentration based on cell-type-specific reference measurements*

Description

Main function "DeconRNASeq" implements an nonnegative decomposition by quadratic programming as datasets = signature*A, where "datasets" are the originally measured data matrix (e.g. genes by samples), "signature" is the signature matrix (genes by cell types) and "A" the cell type concentration matrix (cell types by samples)

Details

Package:	DeconRNASeq
Type:	Package
Version:	1.0
Date:	2012-05-25
License:	GPL version 2 or later

DeconRNASeq(datasets, signature)

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

References

Gong, T., et al. (2011) Optimal Deconvolution of Transcriptional Profiling Data Using Quadratic Programming with Application to Complex Clinical Blood Samples, PLoS One, 6, e27156.

condplot *Draw the plot of the condition numbers vs. the number of genes in the signature matrix*

Description

A function is used to draw the plot of the condition number of signature matrices of all sizes, from a handful of genes in one extreme to the whole signature in the other.

Usage

condplot(step, cond)

Arguments

step	an array with the number of genes used to calculate the condition numbers of signature matrices, default stepwise = 20
cond	an array with the condition numbers of signature matrices

Value

a plot for the condition numbers of signature matrices

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

References

Gong, T., et al. (2011) Optimal Deconvolution of Transcriptional Profiling Data Using Quadratic Programming with Application to Complex Clinical Blood Samples, PLoS One, 6, e27156.

Examples

```
library(DeconRNASeq)
##############################################################
## toy data example:


step <- seq(20,1000, by=20) #every 20 genes
## cell type-specific gene expression matrix:
x.signature <- matrix(rexp(2000),ncol=2)
sig.cond <- sapply(step, function(x) kappa(scale(x.signature[1:x,])))
function (step, cond)
```

---

datasets *data objects for liver and kidney mixing samples*

Description

A data drame providing the RPKM of seven mixing samples.

Usage

```
datasets
```

Format

A data frame with 31979 genes' expression on the 7 mixing samples: reads.1, reads.2, reads.3, reads.4, reads.5, reads.6, reads.7

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

```
data(liver_kidney)
```

DeconRNASeq *Function for Deconvolution of Complex Samples from RNA-Seq.*

Description

This function predicts proportions of constituting cell types from gene expression data generated from RNA-Seq data. Perform nonnegative quadratic programming to get per-sample based globally optimized solutions for constituting cell types .

Usage

DeconRNASeq(datasets, signatures, proportions = NULL, checksig = FALSE, known.prop = FALSE)

Arguments

datasets	measured mixture data matrix, genes (transcripts) e.g. gene counts by samples, . The user can choose the appropriate counts, RPKM, FPKM etc..
signatures	signature matrix from different tissue/cell types, genes (transcripts) by cell types. For gene counts, the user can choose the appropriate counts, RPKM, FPKM etc..
proportions	proportion matrix from different tissue/cell types.
checksig	whether the condition number of signature matrix should be checked
known.prop	whether the proportions of cell types have been known in advanced for proof of concept

Details

Data in the originally measured mixuture sample matrix: datasets and reference matrix: signatures, need to be non-negative. We recommend to deconvolute without log-scale.

Value

Function DeconRNA-Seq returns a list of results

out.all	estimated cell type fraction matrix for all the mixture samples
out.pca	svd calculated PCA on the mixture samples to estimate the number of pure sources according to the cumulative R2

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

References

Gong, T., et al. (2011) Optimal Deconvolution of Transcriptional Profiling Data Using Quadratic Programming with Application to Complex Clinical Blood Samples, PLoS One, 6, e27156.

Examples

```
##source("DeconRNASeq.R")
### multi_tissue: expression profiles for 10 mixing samples from multiple tissues
#data(multi_tissue.rda)

#datasets <- x.data[,2:11]
#signatures <- x.signature.filtered.optimal[,2:6]
#proportions <- fraction

#DeconRNASeq(datasets, signatures, proportions, checksig=FALSE, known.prop = TRUE)
#
```

---

fraction *mixing fractions for multi-tissues mixing samples*

Description

A data frame providing the fractions from 5 tissues in the mixing samples

Usage

```
fraction
```

Format

A martix whose rows are mixing samples' name and columns are fractions from pure tissues including brain, muscle, lung, liver and heart

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

```
data(multi_tissue)
```

---

liver_kidney *data objects for liver and kidney mixing samples*

Description

a list containing:

1) datasets:a data drame providing the RPKM of seven mixing samples.

2) proportions: a data frame providing the fractions for liver and kidney in the mixing samples

3) signatures: a data frame providing the expression values from pure liver and kidney samples

Usage

```
liver_kidney
```

Format

A list 1) a data frame with 31979 genes' expression on the 7 mixing samples: reads.1, reads.2, reads.3, reads.4, reads.5, reads.6, reads.7

2) a martix whose rows are mixing samples' name and columns are fractions from pure live and kidney tissues

3) a data matrix with 630 expressions from pure liver and kidney tissues

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

```
data(liver_kidney)
```

---

multiplot *Draw the plots of proportions of cells determined from deconvolution vs. proportions of the cells actually mixed (when available) with RMSE.*

Description

A function is used to draw the multiple plots of proportions of cells determined from deconvolution vs. proportions of the cells actually mixed. Each plot corresponds to one tissue/cell.

Usage

```
multiplot(..., plotlist = NULL, cols)
```

Arguments

...	any number of the plot objects that store the scatter plots for all the cells/tissue types
plotlist	any other plot objects
cols	columns of the plots, default = 1

Value

A pdf file with the plots of proportions of cells determined from deconvolution vs. proportions of the cells actually mixed with RMSE

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

References

Gong, T., et al. (2011) Optimal Deconvolution of Transcriptional Profiling Data Using Quadratic Programming with Application to Complex Clinical Blood Samples, PLoS One, 6, e27156.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
function (..., plotlist = NULL, cols)
{
    pdf("scatterplots.pdf")
    require(grid)
    plots <- c(list(...), plotlist)
    numPlots = length(plots)
    plotCols = cols
    plotRows = ceiling(numPlots/plotCols)
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(plotRows, plotCols)))
    vplayout <- function(x, y) viewport(layout.pos.row = x, layout.pos.col = y)
    for (i in 1:numPlots) {
        curRow = ceiling(i/plotCols)
        curCol = (i - 1)%%plotCols + 1
        print(plots[[i]], vp = vplayout(curRow, curCol))
    }
    dev.off()
  }
```

---

multi_tissue *data objects for multi-tissues mixing samples*

Description

a list containing:

1) x.data:a data drame providing the RPKM of nine mixing samples.

2) x.signatures: a data frame providing the expression values from pure brain, muscle, lung, liver and heart samples.

3) x.signatures.filtered: a data frame providing the expression values from pure brain, muscle, lung, liver and heart samples after filtering.

4) x.signatures.filtered.optimal: a data frame providing the expression values from pure brain, muscle, lung, liver and heart samples used for the example in DeconRNA-Seq.

5)fraction: a data frame providing the fractions from 5 tissues in the mixing samples

Usage

```
multi_tissue
```

Format

A list 1) a matrix with all the genes' expression in the mixing samples: the first two columns are corresponding to the RefSeq accession numbers and gene symbols

2) a martix whose rows are gene symbols and columns are RPKM expressions from pure tissues.

3) a martix whose rows are gene symbols and columns are RPKM expressions from pure tissues: the genes with RPKM less than 200 within any of the five tissues have been filtered.

4) a martix whose rows are gene symbols and columns are RPKM expressions from pure tissues: based on the filtered signature matrix, the optimal number of genes have been selected for the deconvolution according to the condition numbers

5) a martix whose rows are mixing samples' name and columns are fractions from pure tissues including brain, muscle, lung, liver and heart

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

```
data(multi_tissue)
```

---

proportions *proportions for liver and kidney mixing samples*

Description

proportions: a data frame providing the fractions for liver and kidney in the mixing samples

Usage

```
proportions
```

Format

a martix whose rows are mixing samples' name and columns are fractions from pure live and kidney tissues

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

```
data(liver_kidney)
```

rmse *Calculate the differences between proportions predicted by deconvolution and the values actually measured*

Description

A function is used to calculate the root-mean-square error (RMSE) for the accurracy of estimated proportions.

Usage

```
rmse(x, y)
```

Arguments

x	proportions from the actual measurement
y	estimated proportions from our deconvolution

Value

A number for RMSE

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

References

Gong, T., et al. (2011) Optimal Deconvolution of Transcriptional Profiling Data Using Quadratic Programming with Application to Complex Clinical Blood Samples, PLoS One, 6, e27156.

signatures *data objects for liver and kidney pure samples*

Description

signatures: a data frame providing the expression values from pure liver and kidney samples

Usage

```
signatures
```

Format

a data matrix with 630 expressions from pure liver and kidney tissues

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

data(liver_kidney)

---

x.data *data objects for multi-tissues mixing samples*

Description

A data drame providing the RPKM of nine mixing samples.

Usage

x.data

Format

A matrix with all the genes' expression in the mixing samples: the first two columns are corresponding to the RefSeq accession numbers and gene symbols

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

data(multi_tissue)

---

x.signature *data objects for multi-tissues pure samples*

Description

A data frame providing the expression values from pure brain, muscle, lung, liver and heart samples.

Usage

x.signature

Format

A martix whose rows are gene symbols and columns are RPKM expressions from pure tissues.

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

data(multi_tissue)

---

x.signature.filtered *filtered signatures for multi-tissues pure samples*

Description

A data frame providing the expression values from pure brain, muscle, lung, liver and heart samples after filtering.

Usage

```
x.signature.filtered
```

Format

A martix whose rows are gene symbols and columns are RPKM expressions from pure tissues: the genes with RPKM less than 200 within any of the five tissues have been filtered.

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

```
data(multi_tissue)
```

---

x.signature.filtered.optimal
selected signatures from multi-tissues pure samples

Description

A data frame providing the expression values from pure brain, muscle, lung, liver and heart samples used for the example in DeconRNA-Seq.

Usage

```
x.signature.filtered.optimal
```

Format

A martix whose rows are gene symbols and columns are RPKM expressions from pure tissues: based on the filtered signature matrix, the optimal number of genes have been selected for the deconvolution according to the condition numbers

Author(s)

Ting Gong <tinggong@gmail.com> Joseph D. Szustakowski <joseph.szustakowski@novartis.com>

Examples

```
data(multi_tissue)
```

Index