

Package ‘estimate’

October 4, 2016

Version 1.0.13

Date 2016-09-26

Title Estimate of Stromal and Immune Cells in Malignant Tumor Tissues
from Expression Data

Author Kosuke Yoshihara, Hoon Kim, Roel GW Verhaak

Maintainer Kosuke Yoshihara <kyoshihara@mdanderson.org>

Description Predicts tumor purity, and the presence of stromal/immune cells in tumor tissues using gene expression data.

Depends R (>= 2.15.1)

Imports utils, grDevices, graphics

License GPL-2

LazyData yes

NeedsCompilation no

URL <https://r-forge.r-project.org/projects/estimate/>

Repository R-Forge

Repository/R-Forge/Project estimate

Repository/R-Forge/Revision 21

Repository/R-Forge/DateTimeStamp 2016-10-04 21:13:08

Date/Publication 2016-10-04 21:13:08

R topics documented:

estimate-package	2
common_genes-data	3
estimateScore	4
filterCommonGenes	5
outputGCT	6
plotPurity	7
PurityDataAffy-data	8
SI_geneset-data	9

Index	10
--------------	-----------

estimate-package

ESTIMATE

Description

ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) is a package for predicting tumor purity, and the presence of stromal/immune cells in tumor tissues using gene expression data.

Details

Package: estimate
Type: Package
Version: 1.0.13
Date: 2016-09-26
License: GPL-2

For a complete list of functions, use `library(help="estimate")`.

Author(s)

Kosuke Yoshihara <kyoshihara@mdanderson.org>, P. Roebuck <proebuck@mdanderson.org>

References

Subramanian A, et al.

"Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles."

Proc Natl Acad Sci U S A 2005, 102:15545-15550.

Barbie DA, et al.

"Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1."

Nature 2009, 462:108-112.

Verhaak RG, et al.

"Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1."

Cancer Cell 2010, 17:98-110.

common_genes-data	<i>10,412 common genes</i>
-------------------	----------------------------

Description

This data set contains the information on 10,412 genes that are common among six platforms:

- Affymetrix HG-U133Plus2.0
- Affymetrix HT-HG-U133A
- Affymetrix Human X3P
- Agilent 4x44K (G4112F)
- Agilent G4502A
- Illumina HiSeq RNA sequence

All gene information (according to NCBI) was current as of this package's release date.

Usage

```
data(common_genes)
```

Format

The object `common_genes` is a `data.frame` with components:

EntrezID integer scalar specifying Entrez Gene ID

GeneSymbol character string specifying official HUGO gene symbol

Synonyms character string specifying alternative gene symbols, each delimited by vertical bar

GeneName character string specifying official HUGO gene name

Chromosome character string specifying chromosomal location

Author(s)

Kosuke Yoshihara <kyoshihara@mdanderson.org>

estimateScore	<i>Calculation of stromal, immune, and ESTIMATE scores</i>
---------------	--

Description

This function computes stromal, immune, and ESTIMATE scores per sample using gene expression data.

Usage

```
estimateScore(input.ds,  
              output.ds,  
              platform = c("affymetrix", "agilent", "illumina"))
```

Arguments

input.ds	character string specifying name of input GCT file containing stromal, immune, and estimate scores for each sample
output.ds	character string specifying name of output file
platform	character string indicating platform type. Defaults to "affymetrix"

Details

This method is based on single sample gene set enrichment analysis (ssGSEA) algorithm. This function computes stromal, immune, and ESTIMATE scores using gene-level expression data. For Affymetrix platform data, tumor purity are derived from ESTIMATE scores by applying non-linear squares methods to TCGA Affymetrix expression data (n=995).

Value

Returns data.frame with components:

- StromalScore numeric scalar specifying the presence of stromal cells in tumor tissue
- ImmuneScore numeric scalar specifying the level of infiltrating immune cells in tumor tissue
- ESTIMATEScore numeric scalar specifying tumor cellularity
- TumorPurity numeric scalar specifying ESTIMATE-based tumor purity with value in range[0,1]

Author(s)

Kosuke Yoshihara <kyoshihara@mdanderson.org>

References

Subramanian A, et al.

"Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles."

Proc Natl Acad Sci U S A 2005, 102:15545-15550.

Barbie DA, et al.

"Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1."

Nature 2009, 462:108-112.

Verhaak RG, et al.

"Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1."

Cancer Cell 2010, 17:98-110.

Carter SL, et al.

"Absolute quantification of somatic DNA alterations in human cancer."

Nat Biotechnol 2012, 30:413-421.

Examples

```
in.file <- system.file("extdata", "sample_input.gct", package="estimate")
out.file <- tempfile(pattern="estimate", fileext=".gct")
estimateScore(in.file, out.file)
```

filterCommonGenes

Intersect input data with 10,412 common genes

Description

This function unifies different number of genes per platform against 10,412 common genes.

Usage

```
filterCommonGenes(input.f, output.f, id=c("GeneSymbol", "EntrezID"))
```

Arguments

input.f	character string specifying name of input file, or data.frame including gene-level expression data
output.f	character string specifying name of output file
id	character string indicating which gene identifier to use when matching.

Details

The number of genes in expression data is different for each platform and this difference influences the computational results of stromal and immune scores. To compare stromal, immune and ESTIMATE scores across different platforms or calculate ESTIMATE-based tumor purity using Affymetrix expression data, users need to unify the gene identifiers of the input data against the common genes. Afterwards, the results of the intersection are written to disk in GCT format.

Author(s)

Kosuke Yoshihara <kyoshihara@mdanderson.org>

See Also

[outputGCT](#)

Examples

```
in.file <- system.file("extdata", "sample_input.txt", package="estimate")
out.file <- tempfile(pattern="estimate", fileext=".gct")
filterCommonGenes(in.file, out.file)
```

outputGCT

Write gene expression data in GCT format

Description

This function writes an input file in GCT format.

Usage

```
outputGCT(input.f, output.f)
```

Arguments

input.f	character string specifying name of input file, or data frame including gene-level expression data
output.f	character string specifying name of output file

Details

This function writes gene-level expression data to GCT format file. Input file must be a text format file.

Author(s)

Kosuke Yoshihara <kyoshihara@mdanderson.org>

References

Reich M, et al.
"GenePattern 2.0."
Nature Genetics 2006, 38:500-501.

Examples

```
in.file <- system.file("extdata", "sample_input.txt", package="estimate")
out.file <- tempfile(pattern="estimate", fileext=".gct")
outputGCT(in.file, out.file)
```

plotPurity	<i>Plot tumor purity</i>
------------	--------------------------

Description

Plot tumor purity based on ESTIMATE score.

Usage

```
plotPurity(scores,
           samples="all_samples",
           platform = c("affymetrix", "agilent", "illumina"),
           output.dir="estimated_purity_plots")
```

Arguments

scores	a GCT file output by estimateScore method containing stromal, immune, ESTIMATE scores, and tumor purity per each sample
samples	vector of character strings specifying sample names to be plotted. Defaults to "all_samples", which creates plots for all input samples
platform	character string indicating platform type. Defaults to "affymetrix"
output.dir	character string specifying a directory where plots should be saved. Defaults to "estimated_purity_plots"

Details

This function produces scatterplots (as PNG files) for each requested sample; it plots tumor purity against ESTIMATE score. At present, only the Affymetrix platform is supported.

Author(s)

Kosuke Yoshihara <kyoshihara@mdanderson.org>

Examples

```
in.file <- system.file("extdata", "sample_purity.gct", package="estimate")
plotPurity(scores=in.file)
```

PurityDataAffy-data *Affymetrix data*

Description

This data set contains stromal, immune, and ESTIMATE scores in all Affymetrix expression data (n=995) that was used to develop the formula for predicting tumor purity based on raw estimate score.

This data set also contains tumor purity based on ABSOLUTE algorithm, predicted tumor purity, and 95 ESTIMATE algorithm.

Usage

```
data(PurityDataAffy)
```

Format

The object PurityDataAffy is a data.frame with components:

tumor.purity numeric scalar specifying tumor purity calculated by ABSOLUTE algorithm
StromalScore numeric scalar specifying the presence of stromal cells in tumor tissue
ImmuneScore numeric scalar specifying the level of infiltrating immune cells in tumor tissue
ESTIMATEScore numeric scalar specifying tumor cellularity
fit numeric scalar specifying estimated tumor purity based on ESTIMATE algorithm
lwr.p numeric scalar specifying 5 percent confidence interval
upr.p numeric scalar specifying 95 percent confidence interval

Author(s)

Kosuke Yoshihara <kyoshihara@mdanderson.org>

Source

The Cancer Genome Atlas

References

Carter SL, et al.
"Absolute quantification of somatic DNA alterations in human cancer."
Nat Biotechnol 2012, 30:413-421.

SI_geneset-data *two signatures for estimate*

Description

This data set contains two gene signatures (stromal and immune signatures). The stromal signature including 141 stroma-specific genes is designed to capture the presence of stroma in tumor tissue. The immune signature consisting of 141 immune cell-specific genes represents the infiltration of immune cells in tumor tissue.

Usage

```
data(SI_geneset)
```

Format

The object SI_geneset is a data.frame with components:

StromalSignature character string specifying 141 genes in stromal signature

ImmuneSignature character string specifying 141 genes in immune signature

Note

This data set was based on 'SI_geneset.gmt' (in the 'extdata' directory).

Author(s)

Kosuke Yoshihara <kyoshihara@mdanderson.org>

Index

*Topic **convert**

filterCommonGenes, 5
outputGCT, 6

*Topic **dataset**

common_genes-data, 3
PurityDataAffy-data, 8
SI_geneset-data, 9

*Topic **figure**

plotPurity, 7

*Topic **package**

estimate-package, 2

*Topic **predict**

estimateScore, 4

common_genes (common_genes-data), 3

common_genes-data, 3

estimate (estimate-package), 2

estimate-package, 2

estimateScore, 4

filterCommonGenes, 5

outputGCT, 6, 6

plotPurity, 7

PurityDataAffy (PurityDataAffy-data), 8

PurityDataAffy-data, 8

SI_geneset (SI_geneset-data), 9

SI_geneset-data, 9