Package ‘SCEnt’

July 21, 2025

Title Single Cell Entropy Analysis of Gene Heterogeneity in Cell Populations

Version 0.0.1

Description Analyse single cell RNA sequencing data using entropy to calculate heterogeneity and homogeneity of genes amongst the cell population. From the work of Michael J. Casey, Ruben J. Sanchez-Garcia and Ben D. MacArthur.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.1.2

Imports entropy, tibble

Suggests rmarkdown, knitr, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Hugh Warden [aut, cre] (ORCID: <`https://orcid.org/0000-0002-4308-7316`>)

Maintainer Hugh Warden <`hugh.warden@outlook.com`>

Repository CRAN

Date/Publication 2021-09-30 08:30:04 UTC

Contents

`gene_het` *Find the Heterogeneity of a Gene Within a Population*

Description

Find the Heterogeneity of a Gene Within a Population

Usage

```
gene_het(expr, unit = "log2", normalise = TRUE, transpose = FALSE)
```

Arguments

`expr`	A vector or matrix of gene expressions. For the matrix, genes should be represented as rows and cells as columns.
`unit`	The units to be parsed to the entropy function.
`normalise`	A logical value representing whether the gene frequencies should be normalised into a distribution.
`transpose`	A logical value representing whether the matrix should be transposed before any calculations are performed.

Value

A vector of the information gained from the gene distribution compared to the uniform distribution. The higher the value more heterogeneous the cell is within the population.

Examples

```
#Creating Data
gene1 <- c(0,0,0,0,1,2,3)
gene2 <- c(5,5,3,2,0,0,0)
gene3 <- c(2,0,2,1,3,0,1)
gene4 <- c(3,3,3,3,3,3,3)
gene5 <- c(0,0,0,0,5,0,0)
gene_counts <- matrix(c(gene1,gene2,gene3,gene4,gene5), ncol = 5)
rownames(gene_counts) <- paste0("cell",1:7)
colnames(gene_counts) <- paste0("gene",1:5)

#Calculating Heterogeneity For Each Gene
gene_het(gene1)
gene_het(gene2)
gene_het(gene3)
gene_het(gene4)
gene_het(gene5)

#Calculating Heterogeneity For a Matrix
gene_het(gene_counts)
```

`gene_hom` *Find the Homogeneity of a Gene Within a Population*

Description

Find the Homogeneity of a Gene Within a Population

Usage

```
gene_hom(expr, unit = "log2", normalise = TRUE, transpose = FALSE)
```

Arguments

`expr`	A vector or matrix of gene expressions. For the matrix, genes should be represented as rows and cells as columns.
`unit`	The units to be parsed to the entropy function.
`normalise`	A logical value representing whether the gene frequencies should be normalised into a distribution.
`transpose`	A legical value representing whether the matrix should be transposed before any calculations are performed.

Value

A vector of the information contained in the distribution of each gene. The higher this is, the more homogeneous the gene is within the cell population.

Examples

```
#Creating Data
gene1 <- c(0,0,0,0,1,2,3)
gene2 <- c(5,5,3,2,0,0,0)
gene3 <- c(2,0,2,1,3,0,1)
gene4 <- c(3,3,3,3,3,3,3)
gene5 <- c(0,0,0,0,5,0,0)
gene_counts <- matrix(c(gene1,gene2,gene3,gene4,gene5), ncol = 5)
rownames(gene_counts) <- paste0("cell",1:7)
colnames(gene_counts) <- paste0("gene",1:5)

#Calculating Homogeneity For Each Gene
gene_hom(gene1)
gene_hom(gene2)
gene_hom(gene3)
gene_hom(gene4)
gene_hom(gene5)

#Calculating Homogeneity For a Matrix
gene_hom(gene_counts)
```

`normalise` *Normalise Counts into a Distribution*

Description

A function that takes frequency count data and normalises it into a probability distribution. Only available internally within SCEnt.

Usage

```
normalise(dist)
```

Arguments

`dist`	A vector of a frequency distribution.

Value

A vector of a probability distribution relative to the frequencies.

`scent_select` *Feature Selection by Gene Heterogeneity*

Description

Feature Selection by Gene Heterogeneity

Usage

```
scent_select(
  expr,
  bit_threshold = NULL,
  count_threshold = NULL,
  perc_threshold = NULL,
  unit = "log2",
  normalise = TRUE,
  transpose = FALSE
)
```

Arguments

`expr`	A matrix of gene expression data. Cells should be represented as rows and genes should be represented as columns.
`bit_threshold`	The threshold for the amount of bits of information a gene must add to be selected as a feature. Only one threshold can be used at a time.
`count_threshold`	A number represented how many of the most heterogeneous cells should be selected. Only one threshold can be used at a time.
`perc_threshold`	The percentile of the hetergeneity distribution above which a gene should be to be selected as a feature.
`unit`	The units to be used when calculating entropy.
`normalise`	A logical value representing whether the gene counts should be normalised into a probability distribution.
`transpose`	A logical value representing whether the matrix should be transposed before having any operations computed on it.

Value

A matrix of gene expression values where genes with low heterogeneity have been removed.

Examples

```
#Creating Data
gene1 <- c(0,0,0,0,1,2,3)
gene2 <- c(5,5,3,2,0,0,0)
gene3 <- c(2,0,2,1,3,0,1)
gene4 <- c(3,3,3,3,3,3,3)
gene5 <- c(0,0,0,0,5,0,0)
gene_counts <- matrix(c(gene1,gene2,gene3,gene4,gene5), ncol = 5)
rownames(gene_counts) <- paste0("cell",1:7)
colnames(gene_counts) <- paste0("gene",1:5)

#Performing Feature Selection
scent_select(gene_counts, bit_threshold = 0.85)
scent_select(gene_counts, count_threshold = 2)
scent_select(gene_counts, perc_threshold = 0.25)
```

`scent_select_tidy` *A Tidy Wrapper for Feature Selection by Heterogeneity*

Description

A Tidy Wrapper for Feature Selection by Heterogeneity

Usage

```
scent_select_tidy(
  expr,
  bit_threshold = NULL,
  count_threshold = NULL,
  perc_threshold = NULL,
  unit = "log2",
  normalise = TRUE,
  transpose = FALSE
)
```

Arguments

`expr`	A tibble of gene expression data. Cells should be represented as rows and genes should be represented as columns.
`bit_threshold`	The threshold for the amount of bits of information a gene must add to be selected as a feature. Only one threshold can be used at a time.
`count_threshold`	A number represented how many of the most heterogeneous cells should be selected. Only one threshold can be used at a time.
`perc_threshold`	The percentile of the hetergeneity distribution above which a gene should be to be selected as a feature.
`unit`	The units to be used when calculating entropy.
`normalise`	A logical value representing whether the gene counts should be normalised into a probability distribution.
`transpose`	A logical value representing whether the matrix should be transposed before having any operations computed on it.

Value

A tibble of gene expression values where genes with low heterogeneity have been removed.

Examples

```
#Creating Data
library(tibble)
gene1 <- c(0,0,0,0,1,2,3)
gene2 <- c(5,5,3,2,0,0,0)
gene3 <- c(2,0,2,1,3,0,1)
gene4 <- c(3,3,3,3,3,3,3)
gene5 <- c(0,0,0,0,5,0,0)
gene_counts <- matrix(c(gene1,gene2,gene3,gene4,gene5), ncol = 5)
rownames(gene_counts) <- paste0("cell",1:7)
colnames(gene_counts) <- paste0("gene",1:5)
gene_counts <- as_tibble(gene_counts)

#Performing Feature Selection
scent_select_tidy(gene_counts, bit_threshold = 0.85)
```

```
scent_select_tidy(gene_counts, count_threshold = 2)
scent_select_tidy(gene_counts, perc_threshold = 0.25)
```

Index