

Package: countToFPKM (via r-universe)

August 27, 2024

Title Convert Counts to Fragments per Kilobase of Transcript per Million (FPKM)

Version 1.2.0

Date 2020-07-19

Author Ahmed Alhendi (Dr.)

Maintainer Ahmed Alhendi <asna4@le.ac.uk>

Depends R (>= 3.1.0), ComplexHeatmap, circlize, stats

Description Implements the algorithm described in Trapnell, C. et al. (2010) <[doi:10.1038/nbt.1621](https://doi.org/10.1038/nbt.1621)>. This function takes read counts matrix of RNA-Seq data, feature lengths which can be retrieved using 'biomaRt' package, and the mean fragment lengths which can be calculated using the 'CollectInsertSizeMetrics(Picard)' tool. It then returns a matrix of FPKM normalised data by library size and feature effective length. It also provides the user with a quick and reliable function to generate FPKM heatmap plot of the highly variable features in RNA-Seq dataset.

License GPL-3

URL <https://github.com/AAlhendi1707/countToFPKM>

BugReports <https://github.com/AAlhendi1707/countToFPKM/issues>

Repository <https://aalhendi1707.r-universe.dev>

RemoteUrl <https://github.com/aalhendi1707/counttofpkm>

RemoteRef HEAD

RemoteSha 674bbeb3d042b47a4392f1270bbfca9b833d4b57

Contents

fpkm	2
fpkmheatmap	3
Index	5

fpkm	<i>Convert Counts to Fragments per Kilobase of Transcript per Million (FPKM) Convert counts to Fragments Per Kilobase of transcript per Million mapped reads (FPKM)</i>
------	---

Description

fpkm() function returns a numeric matrix normalized by library size and feature length.

Usage

```
fpkm (counts, featureLength, meanFragmentLength)
```

Arguments

counts	A numeric matrix of raw feature counts
featureLength	A numeric vector with feature lengths which can be obtained using 'biomaRt' package. The length of items should be as the same of rows in read count matrix.
meanFragmentLength	A numeric vector with mean fragment lengths, which can be calculated using 'CollectInsertSizeMetrics(Picard)' tool. The length of items should be as the same of columns in read count matrix.

Details

Implements the algorithm described in Trapnell,C. et al. (2010). "Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation". Nat. Biotechnol., 28, 511-515. doi: 10.1038/nbt.1621. This function takes a matrix of read feature counts of RNA-seq, a numeric vector with feature lengths which can be retrieved using the 'biomaRt' package, and a numeric vector with mean fragment length which can be calculated using the 'CollectInsertSizeMetrics(Picard)' tool. It then returns a matrix of FPKM normalised data by library size and feature effective length. Please see the original manuscript for further details.

Value

A data matrix normalized by library size and feature length.

References

Trapnell,C. et al. (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol., 28, 511-515. doi: 10.1038/nbt.1621.

Lior Pachter. Models for transcript quantification from RNA-Seq. arXiv:1104.3889v2.

Examples

```
library(countToFPKM)

file.readcounts <- system.file("extdata", "RNA-seq.read.counts.csv", package="countToFPKM")
file.annotations <- system.file("extdata", "Biomart.annotations.hg38.txt", package="countToFPKM")
file.sample.metrics <- system.file("extdata", "RNA-seq.samples.metrics.txt", package="countToFPKM")

# Import the read count matrix data into R.
counts <- as.matrix(read.csv(file.readcounts))

# Import feature annotations.
# Assign feature length into a numeric vector.
gene.annotations <- read.table(file.annotations, sep="\t", header=TRUE)
featureLength <- gene.annotations$length

# Import sample metrics.
# Assign mean fragment length into a numeric vector.
samples.metrics <- read.table(file.sample.metrics, sep="\t", header=TRUE)
meanFragmentLength <- samples.metrics$meanFragmentLength

# Return FPKM into a numeric matrix.
fpm_matrix <- fpm (counts, featureLength, meanFragmentLength)
```

fpmheatmap

Convert Counts to Fragments per Kilobase of Transcript per Million (FPKM) A FPKM heatmap plot of the highly variable features in RNA-Seq dataset

Description

fpmheatmap() function returns a heatmap plot of the highly variable features in RNA-Seq dataset.

Usage

```
fpmheatmap(fpm_matrix, topvar=30, showfeaturenames=TRUE, return_log = TRUE)
```

Arguments

fpm_matrix	A data matrix normalized by library size and feature length.
topvar	Number of highly variable features to show in heatmap plot.
showfeaturenames	whether to show the name of features in heatmap plot. The default value is TRUE.
return_log	whether to use log ₁₀ transformation of (fpm+1). The default value is TRUE.

Details

The `fpmheatmap()` function provides users with a robust method to generate a FPKM heatmap plot of the highly variable features in RNA-Seq dataset. It takes an FPKM numeric matrix which can be obtained using `'fpm()'` function as input. By default using Pearson correlation - 1 to measure the distance between features, and Spearman correlation -1 for clustering of samples. By default \log_{10} transformation of $(\text{FPKM}+1)$ is applied to make variation similar across orders of magnitude. It uses the `var()` function to identify the highly variable features. It then uses `Heatmap()` function from the `'ComplexHeatmap'` package to generate a heatmap plot.

Value

A FPKM heatmap plot of the highly variable features in RNA-Seq dataset.

References

Trapnell,C. et al. (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, 28, 511-515. doi: 10.1038/nbt.1621.

Examples

```
library(countToFPKM)

file.readcounts <- system.file("extdata", "RNA-seq.read.counts.csv", package="countToFPKM")
file.annotations <- system.file("extdata", "Biomart.annotations.hg38.txt", package="countToFPKM")
file.sample.metrics <- system.file("extdata", "RNA-seq.samples.metrics.txt", package="countToFPKM")

# Import the read count matrix data into R.
counts <- as.matrix(read.csv(file.readcounts))

# Import feature annotations.
# Assign feature length into a numeric vector.
gene.annotations <- read.table(file.annotations, sep="\t", header=TRUE)
featureLength <- gene.annotations$length

# Import sample metrics.
# Assign mean fragment length into a numeric vector.
samples.metrics <- read.table(file.sample.metrics, sep="\t", header=TRUE)
meanFragmentLength <- samples.metrics$meanFragmentLength

# Return FPKM into a numeric matrix.
fpm_matrix <- fpm (counts, featureLength, meanFragmentLength)

# Plot log10(FPKM+1) heatmap of top 30 highly variable features
fpmheatmap(fpm_matrix, topvar=30, showfeaturenames=TRUE, return_log = TRUE)
```

Index

fpm, 2
fpmheatmap, 3