Package 'countToFPKM'

March 27, 2019

Title Convert counts to Fragments Per Kilobase of transcript per Million (FPKM).

Version 1.0

Date 2019-03-22

Author Ahmed Alhendi (Dr.)

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

Depends R (>= 3.1.0)

Suggests DESeq2, biomaRt

Description Convert a numeric matrix of features with raw feature counts of RNA-seq data to fragments per kilobase of transcript per million mapped reads.

License GPL-3

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

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

R topics documented:

`countToFPKM` *Convert counts to Fragments Per Kilobase of transcript per Million mapped reads (FPKM)*

Description

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

Usage

```
countToFPKM (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. 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 Picard CollectInsertSizeMetrics. 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 RNA-seq, read feature counts data a numeric vector with feature lengths which can be retrieved using biomaRt, and A numeric vector with mean fragment lengths, which can be calculated using Picard CollectInsertSizeMetrics. It then validates the lenght of input data and calculates effective lengths of features in each library to use in normlising the expression of each feature by sample library size and feature effective lenght. 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

```
set.seed(1234)
#Import the read count matrix data into R.
counts <- as.matrix(read.csv("RNA-seq.read.counts.csv"))

#Import feature annotations.
# Assign feature lenght into a numeric vector.
feature.annotations <- read.table("feature.annotations.hg38.txt", sep="\t", header=TRUE)
featureLength <- feature.annotations$length

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

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

Index