

Package ‘tweeDEseqCountData’

September 23, 2012

Title RNA-seq count data employed in the vignette of the tweeDEseq package

Version 1.0.6

Depends Biobase, R (>= 2.10)

Author Juan R Gonzalez <jrgonzalez@creal.cat> and Mikel Esnaola <mesnaola@creal.cat> (with contributions from Robert Castelo <robert.castelo@upf.edu>)

Maintainer Juan R Gonzalez <jrgonzalez@creal.cat>

Description

RNA-seq count data from Pickrell et al. (2010) employed to illustrate the use of the Poisson-Tweedie family of distributions with the tweeDEseq package.

biocViews HighThroughputSequencingData, RNAseqData

License GPL (>=2)

LazyLoad yes

URL <http://www.creal.cat/jrgonzalez/software.htm>

R topics documented:

tweeDEseqCountData-package	2
annotEnsembl63	2
cheung	3
genderGenes	4
hkGenes	5
montgomery	6
pickrell	7

Index	9
--------------	----------

tweedEseqCountData-package

RNA-seq count data employed in the vignette of the tweedEseq package

Description

RNA-seq tables of counts employed in the vignette of the tweedEseq package (Esnaola et al., submitted). These three data sets were downloaded from the ReCount repository at <http://bowtie-bio.sourceforge.net/recount> and contain tables of counts from RNA-seq experiments by Cheung et al. (2010), Montgomery et al. (2010) and Pickrell et al. (2010). The raw RNA-seq data was pre-processed according to the procedures described by Frazee et al. (2011). Please check the individual help pages of each data set for further details.

Source

V.G. Cheung, R.R. Nayak, I.X. Wang, S. Elwyn, S.M. Cousins, M. Morley and R.S. Spielman. *Plos Biology*, 8(9), pii:e1000480, 2010.

A.C. Frazee, B. Langmead and J.T. Leek. ReCount: a multi-experiment resource of analysis-ready RNA-seq gene count datasets. *BMC Bioinformatics*, 12:449, 2011.

S.B. Montgomery, M. Sammeth, M. Gutierrez-Arcelus, R.P. Lach, C. Ingle, J. Nisbett, R. Guigo and E.T. Dermitzakis. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*, 464:773-777, 2010.

J.K. Pickrell, J.C. Marioni, A.A. Pai, J.F. Degner, B.E. Engelhardt, E. Nkadori, J.B. Veyrieras, M. Stephens, Y. Gilad, and J.K. Pritchard. *Nature*, 464:768-772, 2010.

References

M. Esnaola, P. Puig, D. Gonzalez, R. Castelo, J.R. Gonzalez. Gene-specific count data distributions are required in RNA-seq experiments with extensive replication, *submitted*.

See Also

[annotEnsembl63](#) [cheung](#) [montgomery](#) [pickrell](#) [genderGenes](#) [hkGenes](#)

annotEnsembl63

Annotation data from Ensemble Release 63

Description

Annotation data for the human genes forming the tables of counts in [pickrell.eset](#), [montgomery.eset](#) and [cheung.eset](#).

Usage

```
data(annotEnsembl63)
```

Format

Symbol: gene symbol according to the HUGO Gene Nomenclature Committee (HGNC). Chr: chromosome. Start: start chromosomal position (Human genome version GRCh37). End: end chromosomal position (Human genome version GRCh37). EntrezID: Entrez gene identifier. Description: Short description of the gene. Length: Length of the longest cDNA of this gene. GCcontent: G+C content of the longest cDNA of this gene.

Details

Data for all columns except Length and GCcontent was retrieved from Ensembl release 63 using the biomaRt package. Data in columns Length and GCcontent was obtained by downloading the set of Ensembl Release 63 human cDNAs at ftp://ftp.ensembl.org/pub/release-63/fasta/homo_sapiens/cdna/Homo_sapiens.GRCh37.63.cdna.all.fa.gz and selecting the longest cDNA for each Ensembl Gene from which length and G+C content was calculated.

References

M. Esnaola, P. Puig, D. Gonzalez, R. Castelo, and J.R. Gonzalez. Gene-specific count data distributions are required in RNA-seq experiments with extensive replication. Submitted.

See Also

[pickrell](#) [pickrellNorm](#) [montgomery](#) [genderGenes](#) [hkGenes](#)

Examples

```
data(annotEnsembl63)
dim(annotEnsembl63)
head(annotEnsembl63)
```

cheung

RNA-seq count data from Cheung et al. (2010)

Description

ExpressionSet object containing RNA-seq count data from lymphoblastoid cell lines from 41 unrelated Caucasian individuals of European descent. These count data are employed in the vignette of the package `tweedEseq` Esnaola et al. (submitted). The original experimental data was published by Cheung et al. (2010) and the table of counts in this ExpressionSet object corresponds to the one in the ReCount repository available at <http://bowtie-bio.sourceforge.net/recount>. Details on the pre-processing steps to obtain this table of counts from the raw reads of Cheung et al. (2010) are provided on that website and in the publication by Frazee et al. (2011).

Usage

```
data(cheung)
```

Format

`cheung.eset`: ExpressionSet object containing read counts for 52,580 Ensembl genes for each of the 41 Caucasian individuals of European descent.

Details

The table of counts is stored in the `AssayData` slot of an `ExpressionSet` object called `cheung.eset` whose phenotypic data contains the gender of each individual, among other bits of information.

Source

V.G. Cheung, R.R. Nayak, I.X. Wang, S. Elwyn, S.M. Cousins, M. Morley and R.S. Spielman. *Plos Biology*, 8(9), pii:e1000480, 2010.

A.C. Frazee, B. Langmead and J.T. Leek. ReCount: a multi-experiment resource of analysis-ready RNA-seq gene count datasets. *BMC Bioinformatics*, 12:449, 2011.

References

M. Esnaola, P. Puig, D. Gonzalez, R. Castelo, J.R. Gonzalez. Gene-specific count data distributions are required in RNA-seq experiments with extensive replication. Submitted.

See Also

[annotEnsembl63](#) [pickrell](#) [cheung](#) [genderGenes](#) [hkGenes](#)

Examples

```
suppressMessages(library(Biobase))
data(cheung)
cheung.eset
table(cheung.eset$gender)
```

genderGenes

Genes with documented sex-specific expression

Description

Genes with documented sex-specific expression and occurring within the set of genes that form the tables of counts in [pickrell.eset](#), [montgomery.eset](#) and [cheung.eset](#).

Usage

```
data(genderGenes)
```

Format

`msYgenes`: Ensembl gene identifiers from genes belonging to the male-specific region of chromosome Y (Skaletsky et al., 2003).

`XiEgenes`: Ensembl gene identifiers from genes located in the X chromosome and which have been reported to escape X-inactivation.

Details

These two lists of genes form a gold-standard set of genes with documented sex-specific expression which have been employed in the assessment of the method for differential expression analysis implemented in the `tweeDEseq` package (Esnaola et al., *submitted*). Both gene lists are restricted to genes occurring within the set of genes that form the tables of counts in [pickrell.eset](#), [montgomery.eset](#) and [cheung.eset](#).

Source

H.S. Skaletsky, T. Kuroda-Kawaguchi, P.J. Minx, H.S. Cordum, L. Hillier, L.G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, K. Delehaunty, H. Du, G. Fewell, L. Fulton, T. Graves, S.F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfling, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S.P. Yang, R.H. Waterston, R.K. Wilson, S. Rozen, and D.C. Page. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423:825–837, 2003.

L. Carrel and H.F. Willard. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, 434:400–404, 2005.

References

M. Esnaola, P. Puig, D. Gonzalez, R. Castelo, J.R. Gonzalez. Gene-specific count data distributions are required in RNA-seq experiments with extensive replication. Submitted.

See Also

[annotEnsembl63](#) [pickrell](#) [pickrellNorm](#) [montgomery](#) [hkGenes](#)

Examples

```
data(genderGenes)
length(msYgenes)
length(XiEgenes)
```

hkGenes

Housekeeping Genes from Eisenberg and Levanon (2003)

Description

Housekeeping genes reported by Eisenberg and Levanon (2003) and occurring within the set of genes that form the tables of counts in [pickrell.eset](#), [montgomery.eset](#) and [cheung.eset](#) in this experimental data package.

Usage

```
data(hkGenes)
```

Format

hkGenes: Ensembl gene identifiers from the list of housekeeping genes occurring within the set of genes that form the tables of counts in [pickrell.eset](#), [montgomery.eset](#) and [cheung.eset](#) in this experimental data package.

Details

This list of genes has been derived from mapping the original list in http://www.cgen.com/supp_info/Housekeeping_genes.html to Ensembl Gene identifiers using the [org.Hs.eg.db](#) package. This list of housekeeping genes has been employed to compare count data distributions from genes with different expression dynamics in (Esnaola et al., *submitted*) and is restricted to genes occurring within the set of genes that form the tables of counts in [pickrell.eset](#), [montgomery.eset](#) and [cheung.eset](#) in this experimental data package.

Source

E. Eisenberg, and E.Y. Levanon. Human housekeeping genes are compact. *Trends Genet*, 19(7):362–365, 2003.

References

M. Esnaola, P. Puig, D. Gonzalez, R. Castelo, and J.R. Gonzalez. Gene-specific count data distributions are required in RNA-seq experiments with extensive replication. Submitted.

See Also

[annotEnsembl63](#) [pickrell](#) [pickrellNorm](#) [montgomery](#) [genderGenes](#)

Examples

```
data(hkGenes)
length(hkGenes)
head(hkGenes)
```

montgomery

RNA-seq count data from Montgomery et al. (2010)

Description

ExpressionSet object containing RNA-seq count data from lymphoblastoid cell lines from 60 unrelated Caucasian individuals of European descent. These count data are employed in the vignette of the package `tweedEseq` Esnaola et al. (submitted). The original experimental data was published by Montgomery et al. (2010) and the table of counts in this ExpressionSet object corresponds to the one in the ReCount repository available at <http://bowtie-bio.sourceforge.net/recount>. Details on the pre-processing steps to obtain this table of counts from the raw reads of Montgomery et al. (2010) are provided on that website and in the publication by Frazee et al. (2011).

Usage

```
data(montgomery)
```

Format

`montgomery.eset`: ExpressionSet object containing read counts for 52,580 Ensembl genes for each of the 60 Caucasian individuals of European descent.

Details

The table of counts is stored in the AssayData slot of an ExpressionSet object called `montgomery.eset` whose phenotypic data contains the gender of each individual, among other bits of information.

Source

S.B. Montgomery, M. Sammeth, M. Gutierrez-Arcelus, R.P. Lach, C. Ingle, J. Nisbett, R. Guigo and E.T. Dermitzakis. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*, 464:773-777, 2010.

A.C. Frazee, B. Langmead and J.T. Leek. ReCount: a multi-experiment resource of analysis-ready RNA-seq gene count datasets. *BMC Bioinformatics*, 12:449, 2011.

References

M. Esnaola, P. Puig, D. Gonzalez, R. Castelo, J.R. Gonzalez. Gene-specific count data distributions are required in RNA-seq experiments with extensive replication. Submitted.

See Also

[annotEnsembl63](#) [pickrell](#) [cheung](#) [genderGenes](#) [hkGenes](#)

Examples

```
suppressMessages(library(Biobase))
data(montgomery)
montgomery.eset
table(montgomery.eset$gender)
```

pickrell

RNA-seq count data from Pickrell et al. (2010)

Description

ExpressionSet objects containing RNA-seq count data from lymphoblastoid cell lines from 69 unrelated Nigerian individuals. These count data are employed in the vignette of the package `tweeDEseq` Esnaola et al. (submitted). The original experimental data was published by Pickrell et al. (2010). The table of counts in `pickrell.eset` corresponds to the one in the ReCount repository available at <http://bowtie-bio.sourceforge.net/recount>. Details on the pre-processing steps to obtain this table of counts from the raw reads of Pickrell et al. (2010) are provided on that website and in the publication by Frazee et al. (2011). The other object `pickrellNorm.eset` contains the corresponding filtered and normalized table of counts.

The table of counts in `pickrell1.eset` was obtained by Esnaola et al. (2010) by the pre-processing steps described on that article and `pickrell1Norm.eset` contains the corresponding filtered and normalized table of counts.

Usage

```
data(pickrell)
```

Format

`pickrell.eset`: ExpressionSet object containing read counts for 52,580 Ensembl genes for each of the 69 Nigerian individuals. `pickrellNorm.eset`: ExpressionSet object containing filtered and normalized read counts for 10,231 Ensembl genes for each of the 69 Nigerian individuals. `pickrell1.eset`: ExpressionSet object containing read counts for 38,415 Ensembl genes for each of the 69 Nigerian individuals. `pickrell1Norm.eset`: ExpressionSet object containing filtered and normalized read counts for 22,060 Ensembl genes for each of the 69 Nigerian individuals.

Details

These tables of counts are stored in the AssayData slot of the previously enumerated ExpressionSet objects whose phenotypic data contains the gender of each individual, among other bits of information. The filtered and normalized table of counts was obtained from the raw counts in `pickrell.eset` and `pickrell1.eset` by first removing genes with less than 0.5 cpm (counts per million reads) in all samples but one and then applying the conditional quantile normalization procedure by Hansen et al. (2011).

Source

J.K. Pickrell, J.C. Marioni, A.A. Pai, J.F. Degner, B.E. Engelhardt, E. Nkadori, J.B. Veyrieras, M. Stephens, Y. Gilad, and J.K. Pritchard *Nature*, 464:768-772, 2010.

A.C. Frazee, B. Langmead and J.T. Leek. ReCount: a multi-experiment resource of analysis-ready RNA-seq gene count datasets. *BMC Bioinformatics*, 12:449, 2011.

References

M. Esnaola, P. Puig, D. Gonzalez, R. Castelo, J.R. Gonzalez. Gene-specific count data distributions are required in RNA-seq experiments with extensive replication. Submitted.

K.D. Hansen, R.A. Irizarry and Z. Wu. Removing technical variability in RNA-seq data using conditional quantile normalization. Johns Hopkins University, Dept. of Biostatistics Working Papers, Paper 227. (<http://www.bepress.com/jhubiostat/paper227>).

See Also

[annotEnsembl63](#) [montgomery cheung genderGenes](#) [hkGenes](#)

Examples

```
suppressMessages(library(Biobase))
data(pickrell)
pickrell.eset
table(pickrell.eset$gender)
data(pickrellNorm)
pickrellNorm.eset
```

Index

*Topic **datasets**

- annotEnsembl63, 2
- cheung, 3
- genderGenes, 4
- hkGenes, 5
- montgomery, 6
- pickrell, 7
- tweeDEseqCountData-package, 2

annotEnsembl63, 2, 2, 4–8

cheung, 2, 3, 4, 7, 8
cheung.eset, 2, 4, 5

genderGenes, 2–4, 4, 6–8

hkGenes, 2–5, 5, 7, 8

montgomery, 2, 3, 5, 6, 6, 8
montgomery.eset, 2, 4, 5
msYgenes (genderGenes), 4

org.Hs.eg.db, 5

pickrell, 2–7, 7
pickrell.eset, 2, 4, 5
pickrell1 (pickrell), 7
pickrell1Norm (pickrell), 7
pickrellNorm, 3, 5, 6
pickrellNorm (pickrell), 7

tweeDEseqCountData-package, 2

XiEgenes (genderGenes), 4