Package ‘geuvPack’

October 12, 2020

Title summarized experiment with expression data from GEUVADIS

Version 1.20.0

Author VJ Carey <stvjc@channing.harvard.edu>

Maintainer VJ Carey <stvjc@channing.harvard.edu>

Description FPKM from GEUVADIS, annotated to gencode regions

Depends R (>= 2.10), SummarizedExperiment

License Artistic-2.0

LazyLoad yes

biocViews ExperimentData, Genome, SequencingData, MicroarrayData, ArrayExpress

git_url https://git.bioconductor.org/packages/geuvPack

git_branch RELEASE_3_11

git_last_commit 55f565f

git_last_commit_date 2020-04-27

Date/Publication 2020-10-12

R topics documented:

geuvPack-package *summarized experiment with expression data from GEUVADIS*

Description

FPKM from GEUVADIS, annotated to gencode regions

could include more things like miRNA read counts

protco and hasout are convenience vectors enumerating protein coding features and features that appear to have outlying values. The outlier condition for gene g is (sd(x)/mad(x)) > 2.

Details

Package:	geuvPack
Version:	0.0.0
Suggests:	
Depends:	GenomicRanges
License:	Private
LazyLoad:	yes
Built:	R 3.1.1; ; 2014-08-22 17:12:50 UTC; unix

gtpath function will get 1000 genomes genotype VCF from Amazon S3; modify if you have local VCF

Author(s)

VJ Carey <stvjc@channing.harvard.edu>

Maintainer: VJ Carey <stvjc@channing.harvard.edu>

`geuFPKM` *Expression data, gene level FPKM, from GEUVADIS*

Description

see `metadata(geuFPKM)`

Usage

```
data("geuFPKM"); data(gs2p)
```

Format

geuFPKM: A RangedSummarizedExperiment object.

gsvs: A result from sva::sva

gs2p: a character vector with gene symbols as values and gencode identifiers as names; gen2sym has gencode tags as values and symbols as names

gencodeV12: a GRanges instance with metadata about V12 gencode genes

Details

FPKM as reported in EBI ArrayExpress E-GEUV-1. Other quantifications may be added in future versions of this package.

"500bffeed8e0f770c157e0189e9e50ae" is the output of digest on the txt.gz file of quantifications from which the assay data in the geuFPKM RangedSummarizedExperiment instance is constructed. This was extracted at Channing Division of Network Medicine on 13 November 2013, and verified to be correct for the contents of the URL below on 11 December 2014.

The README file

`http://www.ebi.ac.uk/arrayexpress/files/E-GEUV-1/GeuvadisRNASeqAnalysisFiles_README.txt`

has the following remarks

Quantification file set:
- Sample set + sample size :
QC-passed: All QC-passed samples including replicates: 660 (mRNA) or 480 (miRNA)
QC-passed unique: Nonredundant set of unique samples used in most analyses: 462 (mRNA); 452 (miRNA)
- Normalization:
None: raw read counts
Library depth: Read counts scaled by total number of mapped reads (mRNA), or total number reads mapping to miRNAs (miRNA) per sample, then adjusted to the median of the sample set (45M for mRNA, 1.2M for miRNA)
Library depth and transcript length: RPKM
Library depth & expressed & PEER: Library depth scaling as above, removal of units with 0 counts in >50

- November 5, 2013 update: The file GD462.GeneQuantRPKM.50FN.samplename.resk10.norm.txt.gz that had the normalization as above PLUS an additional transformation of each gene's values to standard normal has been replaced by GD462.GeneQuantRPKM.50FN.samplename.resk10.txt.gz

Source

ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/GEUV/E-GEUV-1/analysis_results/GD462.GeneQuantRPKM.50FN.samplename.resk10.txt.gz

References

PMID 24037378

Examples

```
data(geuFPKM)
geuFPKM
sd(assay(geuFPKM)[1,])
data(gs2p)
gs2p[1:10] # from rowRanges - use with extractBySymbols in gQTLBase
## Not run:
# this is how the gsvs surrogate variable structure
# was derived
library(sva)
library(geuvPack)
data(geuFPKM)
popm = model.matrix(~popcode, data=colData(geuFPKM))
int = popm[,1,drop=FALSE]
gsvs = sva(assay(geuFPKM), popm, int)
save(gsvs, file="gsvs.rda")

## End(Not run)

## maybe str(geuFPKM) ; plot(geuFPKM) ...
```

gsvs *output of SVA on geuvadis FPKM*

Description

output of SVA on geuvadis FPKM

Usage

```
data("gsvs")
```

Format

The format is: List of 4 $ sv : num [1:462, 1:6] -0.01355 0.00506 0.5762 -0.00857 -0.01388 ... $ pprob.gam: num [1:23722] 0.741 0.996 0.829 0.754 1 ... $ pprob.b : num [1:23722] 0.0599 0.9921 0.3887 0.4483 0.8637 ... $ n.sv : num 6

Details

see the example for information on generating the sv

Examples

```
data(geuFPKM)
geuFPKM
sd(assay(geuFPKM)[1,])
data(gs2p)
gs2p[1:10] # from rowRanges - use with extractBySymbols in gQTLBase
## Not run:
# this is how the gsvs surrogate variable structure
# was derived
 library(sva)
 library(geuvPack)
 data(geuFPKM)
 popm = model.matrix(~popcode, data=colData(geuFPKM))
 int = popm[,1,drop=FALSE]
 gsvs = sva(assay(geuFPKM), popm, int)
 save(gsvs, file="gsvs.rda")

## End(Not run)
```

gtpath *generate path for a VCF file for 1000 genomes genotypes*

Description

generate path for a VCF file for 1000 genomes genotypes

Usage

```
gtpath(chrdigit, useS3=TRUE, tmplate)
```

Arguments

`chrdigit` suffix to 'chr' to get the chromosome id

`useS3` logical, if TRUE, returns URL of an Amazon S3 bucket-resident 1000 genomes VCF

`tmplate` character string, used only if `useS3` is `FALSE`. The substring `"%%N%%"` will be replaced by value of `chrdigit`

Details

A subset of samples in GEUVADIS have genotypes recorded in 1000 genomes. This function creates the URL for the VCF corresponding to a chromosome. `cisAssoc` can operate on these genotypes.

Value

character string

Examples

```
gtpath(1)
gtpath(1, FALSE, "/tmp/my%%N%%.vcf.gz")
```

Index