Package 'geuvPack'

October 7, 2015

Title summarized experiment with expression data from GEUVADIS

Version 1.0.0

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

Description FPKM from GEUVADIS, annotated to gencode regions

Suggests

Depends GenomicRanges

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

License Artistic-2.0

LazyLoad yes

biocViews ExperimentData, Genome, SequencingData, MicroarrayData, ArrayExpress

NeedsCompilation no

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

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 the exptData

Usage

```
data("geuFPKM")
```

Format

The format is: Formal class 'SummarizedExperiment' [package "GenomicRanges"] with 4 slots ..@ exptData:Formal class 'SimpleList' [package "IRanges"] with 4 slots .. .. ..@ listData :List of 2 .. ..$ MIAME :Formal class 'MIAME' [package "Biobase"] with 13 slots .. ..@ name : chr "Lappalainen T" .. ..@ lab : chr NA .. ..@ contact : chr "" .. ..@ title : chr "Transcriptome and genome sequencing uncovers functional variation in humans." .. ..@ abstract : chr "Genome sequencing projects are discovering millions of genetic variants in humans, and interpretation of their functional effec"| __truncated__ .. ..@ url : chr "" .. ..@ pubMedIds : chr "24037378" .. ..@ samples : list() .. ..@ hybridizations : list() .. ..@ normControls : list() .. ..@ preprocessing : list() .. ..@ other : list() .. ..@ .__classVersion__:Formal class 'Versions' [package "Biobase"] with 1 slots .. ..@ .Data:List of 2 .. ..$: int [1:3] 1 0 0 .. ..$: int [1:3] 1 1 0 .. ..$ constrHist: chr [1:79] "x = read.delim(\"GD462.GeneQuantRPKM.50FN.samplename.resk10.txt.gz\", sep=\"\t\", h=TRHE)" "x = read.delim(\"GD462.GeneQuantRPKM.50FN.samplename.resk10.txt.gz\", sep=\"\t\", h=TRUE)" "x = read.delim(\"GD462.GeneQuantRPKM.50FN.samplename.resk10.txt.gz\", sep=\"\t\", h=TRUE, nrow=5)" "x" ... .@ elementType : chr "ANY" .. .. ..@ elementMetadata: NULL

.. .. ..@ metadata : list() ..@ rowData :Formal class 'GRanges' [package "GenomicRanges"] with 6 slots .. .. ..@ seqnames :Formal class 'Rle' [package "S4Vectors"] with 4 slots .. ..@ values : Factor w/ 25 levels "chr1","chr2",..: 5 7 12 10 23 7 15 23 11 10 ... .@ lengths : int [1:22493] 1 1 1 1 1 1 1 1 1 1 ... .@ elementMetadata: NULL .. ..@ metadata : list() .. .. ..@ ranges :Formal class 'IRanges' [package "IRanges"] with 6 slots .. ..@ start : int [1:23722] 59783540 48128225 57846106 116054583 131157293 22157909 23096869 134944381 1781578 116449902 ... .@ width : int [1:23722] 59945 20106 6958 109933 52679 238855 8464 9614 2139 492 ... .@ NAMES : chr [1:23722] "ENSG00000152931.6" "ENSG00000183696.9" "ENSG00000139269.2" "ENSG00000169129.8" ... .@ elementType : chr "integer" .. ..@ elementMetadata: NULL .. ..@ metadata : list() .. .. ..@ strand :Formal class 'Rle' [package "S4Vectors"] with 4 slots .. ..@ values : Factor w/ 3 levels "+","-","*": 1 2 1 2 1 2 1 2 1 2 ... .@ lengths : int [1:11852] 3 1 1 1 1 1 1 1 3 4 ... .@ elementMetadata: NULL .. ..@ metadata : list() .. .. ..@ elementMetadata:Formal class 'DataFrame' [package "IRanges"] with 6 slots .. ..@ rownames : NULL .. ..@ nrows : int 23722 .. ..@ listData :List of 18 .. ..$ source : Factor w/ 2 levels "ENSEMBL","HAVANA": 2 2 2 2 2 2 2 2 2 2 ... .$ type : Factor w/ 8 levels "CDS","exon","gene",..: 3 3 3 3 3 3 3 3 3 3 ... .$ score : num [1:23722] NA NA NA NA NA NA NA NA NA NA ... .$ phase : int [1:23722] NA NA NA NA NA NA NA NA NA NA ... .$ gene_id : chr [1:23722] "ENSG00000152931.6" "ENSG00000183696.9" "ENSG00000139269.2" "ENSG00000169129.8" ... .$ transcript_id : chr [1:23722] "ENSG00000152931.6" "ENSG00000183696.9" "ENSG00000139269.2" "ENSG00000169129.8" ... .$ gene_type : chr [1:23722] "lincRNA" "protein_coding" "protein_coding" "protein_coding" ... .$ gene_status : chr [1:23722] "KNOWN" "KNOWN" "KNOWN" "KNOWN" ... .$ gene_name : chr [1:23722] "PART1" "UPP1" "INHBE" "AFAP1L2" ... .$ transcript_type : chr [1:23722] "lincRNA" "protein_coding" "protein_coding" "protein_coding" ... .$ transcript_status: chr [1:23722] "KNOWN" "KNOWN" "KNOWN" "KNOWN" ... .$ transcript_name : chr [1:23722] "PART1" "UPP1" "INHBE" "AFAP1L2" ... .$ level : num [1:23722] 2 2 2 2 1 2 1 2 2 1 ... .$ havana_gene : chr [1:23722] "OTTHUMG00000162213.2" "OTTHUMG00000129253.2" "OTTHUMG00000169995.1" "OTTHUMG00000019086.3" ... .$ havana_transcript: chr [1:23722] NA NA NA NA ... .$ ont : chr [1:23722] NA NA NA NA ... .$ tag : chr [1:23722] NA NA NA NA ... .$ ccdsid : chr [1:23722] NA NA NA NA ... .@ elementType : chr "ANY" .. ..@ elementMetadata: NULL .. ..@ metadata : list() .. .. ..@ seqinfo :Formal class 'Seqinfo' [package "GenomeInfoDb"] with 4 slots .. ..@ seqnames : chr [1:25] "chr1" "chr2" "chr3" "chr4" ... .@ seqlengths : int [1:25] 249250621 243199373 198022430 191154276 180915260 171115067 159138663 146364022 141213431 135534747 ... .@ is_circular: logi [1:25] FALSE FALSE FALSE FALSE FALSE FALSE ... .@ genome : chr [1:25] "hg19" "hg19" "hg19" "hg19" ... .@ metadata : list() ..@ colData :Formal class 'DataFrame' [package "IRanges"] with 6 slots .. .. ..@ rownames : chr [1:462] "HG00096" "HG00097" "HG00099" "HG00100" ... .@ nrows : int 462 .. .. ..@ listData : Named list() .. .. ..@ elementType : chr "ANY" .. .. ..@ elementMetadata: NULL .. .. ..@ metadata : list() ..@ assays :Reference class 'ShallowSimpleListAssays' [package "GenomicRanges"] with 1 fields .. ..$ data:Formal class 'SimpleList' [package "IRanges"] with 4 slots .. ..@ listData :List of 1 .. ..$ exprs: num [1:23722, 1:462] 0.102 8.184 1.2 0.832 27.646 ... .- attr(*, "dimnames")=List of 2 .. ..$: chr [1:23722] "ENSG00000152931.6" "ENSG00000183696.9" "ENSG00000139269.2" "ENSG00000169129.8" ... .$: chr [1:462] "HG00096" "HG00097" "HG00099" "HG00100" ... .@ elementType : chr "ANY" .. ..@ elementMetadata: NULL .. ..@ metadata : list() .. ..and

12 methods,

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 SummarizedExperiment 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,])
```

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

---

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