Package 'cnvGSAdata'

September 23, 2012

Title Data used in the vignette of the cnvGSA package

Version 0.99.2

Author Robert Ziman <rziman@gmail.com>

Description This package contains the data used in the vignette of the cnvGSA package.

Maintainer Robert Ziman <rziman@gmail.com>

License LGPL

biocViews ExperimentData

Depends R (>= 2.10), cnvGSA

R topics documented:

cnvGSAdata-package 1
cnv.gvf 2
cnvGSA_input_example.RData 2
cnvGSA_output_example.RData 3
gsData.gmt 3
merge_00k_flank_hg18_refGene_jun_2011_exon.gff 4
merge_00k_flank_hg18_refGene_jun_2011_transcript.gff 5
s2class.txt 6

Index **7**

`cnvGSAdata-package` *Data used in the examples and vignette of the cnvGSA package*

Description

This package contains data used in the examples and vignette of the cnvGSA package.

Details

Raw input files: `cnv.gvf` - CNV data `gsData.gmt` - Gene-set data `merge_00k_flank_hg18_refGene_jun_2011_exon.g` - Gene map for exons `merge_00k_flank_hg18_refGene_jun_2011_transcript.gff` - Gene map for transcripts `s2class.txt` - Sample-to-class information (for the CNV data)

Pre-built input: `cnvGSA_input_example.RData`

Saved output: `cnvGSA_output_example.RData`

Author(s)

Robert Ziman <rziman@gmail.com>

---

cnv.gvf *Rare CNV data from Pinto et al. 2010 ASD study*

Description

The file contains rare CNV data similar to that used in the Pinto et al. 2010 ASD study.

Details

The data is stored in the Genome Variation Format http://www.sequenceontology.org/resources/gvf.html.

Source

Pinto, D et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature.* 2010 Jul 15; 466(7304): 368–72.

For specific code and details on how the gene-sets were compiled, see http://baderlab.org/GeneSetDB_02

Examples

```
library( "cnvGSA" )
cnvFile <- system.file( "extdata", "cnv.gvf", package="cnvGSAdata" )
cnv <- readGVF( cnvFile )
```

---

cnvGSA_input_example.RData
cnvGSA example input

Description

Pre-built input used in the full workflow example in the cnvGSA vignette.

Usage

```
data("cnvGSA_input_example")
```

Details

The dataset contains input, a single object of class CnvGSAInput as defined in the cnvGSA package. When processed by cnvGSAFisher() – the main function in that package – it produces as its output an object of class CnvGSAOutput (such as the one stored in cnvGSA_output_example.RData).

cnvGSA_output_example.RData *cnvGSA example output*

Description

Saved output from the full workflow example in the cnvGSA vignette.

Usage

```
data("cnvGSA_output_example")
```

Details

The dataset contains output, a single object of class CnvGSAOutput as defined in the cnvGSA package. CnvGSAOutput is a simple S4 class containing a slot for each data structure output by cnvGSAFisher (the main function in the package):

cnvData

enrRes

burdenSample

burdenGs

geneData

cnvData contains the original and filtered CNV data, enrRes contains the gene-set enrichment results, and burdenSample, burdenGs, and geneData contain burden analysis and gene-centric statistics that can be used to ensure the validity of the enrichment results.

See the cnvGSA vignette for complete details and discussion.

gsData.gmt *Gene-set data from Pinto et al. 2010 ASD study*

Description

The file contains gene-set data similar to that used in the Pinto et al. 2010 ASD study.

Details

The data is stored in the Gene Matrix Transposed format http://www.broadinstitute.org/cancer/software/gsea/wiki/index.php/Dataformats.

Source

Pinto, D et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature.* 2010 Jul 15; 466(7304): 368–72.

Examples

```
library( "cnvGSA" )
gsDataFile <- system.file( "extdata", "gsData.gmt", package="cnvGSAdata" )
gsData <- readGMT( gsDataFile )
```

merge_00k_flank_hg18_refGene_jun_2011_exon.gff *Gene map for exons in the rare CNV association test of the Pinto et al. 2010 ASD study*

Description

Contains exon coordinates similar to that used in Pinto et al. 2010 to determine genes affected by the rare CNVs identified in that study.

Details

This file can be used to create the genemap data frame required by the getCnvGenes function of the main cnvGSA package. For more information on the GFF format, see its specification: (http://www.sanger.ac.uk/resources/software/gff/spec.html)

Source

Pinto, D et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature.* 2010 Jul 15; 466(7304): 368–72.

Examples

```
genemapFile <- system.file(
"extdata",
"merge_00k_flank_hg18_refGene_jun_2011_exon.gff",
package = "cnvGSAdata"
)
fields <- read.table (
genemapFile,
sep = "\t",
comment.char = "",
quote = "\"",
header = FALSE,
stringsAsFactors = FALSE
)
genemap_exon <- data.frame(
Chr = fields[,1],
Coord_i = fields[,4],
Coord_f = fields[,5],
GeneID = fields[,11],
stringsAsFactors = FALSE
)
genemap_exon$Chr <- sub( genemap_exon$Chr, pattern = "chr", replacement = "" )
```

merge_00k_flank_hg18_refGene_jun_2011_transcript.gff
Gene map for transcripts in the rare CNV association test of the Pinto et al. 2010 ASD study

Description

Contains transcript coordinates similar to that used in Pinto et al. 2010 to determine genes affected by the rare CNVs identified in that study.

Details

This file can be used to create the genemap data frame required by the getCnvGenes function of the main cnvGSA package. For more information on the GFF format, see its specification: (http://www.sanger.ac.uk/resources/software/gff/spec.html)

Source

Pinto, D et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature.* 2010 Jul 15; 466(7304): 368–72.

Examples

```
genemapFile <- system.file(
"extdata",
"merge_00k_flank_hg18_refGene_jun_2011_transcript.gff",
package = "cnvGSAdata"
)
fields <- read.table (
genemapFile,
sep = "\t",
comment.char = "",
quote = "\"",
header = FALSE,
stringsAsFactors = FALSE
)
genemap_transcript <- data.frame(
Chr = fields[,1],
Coord_i = fields[,4],
Coord_f = fields[,5],
GeneID = fields[,11],
stringsAsFactors = FALSE
)
genemap_transcript$Chr <- sub( genemap_transcript$Chr, pattern = "chr", replacement = "" )
```

s2class.txt *Sample-to-class data from Pinto et al. 2010 ASD study*

Description

The file contains the sample classes similar to that used in the Pinto et al. 2010 ASD study. It should be used in conjuction with the cnv.gvf example file.

Details

The data is stored in a simple tab-delimited format that can be read using read.table().

Source

Pinto, D et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature.* 2010 Jul 15; 466(7304): 368–72.

Examples

```
s2classFile <- system.file( "extdata", "s2class.txt", package="cnvGSAdata" )
s2class <- read.table(
s2classFile,
sep = "\t",
col.names = c("SampleID", "Class"),
stringsAsFactors = FALSE
)
```

Index

cnv.gvf, *1*, 2
cnvGSA_input_example.RData, *1*, 2
cnvGSA_output_example.RData, *1*, 3
cnvGSAdata *(*cnvGSAdata-package*)*, 1
cnvGSAdata-package, 1

gsData.gmt, *1*, 3

input *(*cnvGSA_input_example.RData*)*, 2

merge_00k_flank_hg18_refGene_jun_2011_exon.gff,
        1, 4
merge_00k_flank_hg18_refGene_jun_2011_transcript.gff,
        1, 5

output *(*cnvGSA_output_example.RData*)*, 3

s2class.txt, *1*, 6