Package ‘gwascat’

September 24, 2012

Title representing and modeling data in the NHGRI GWAS catalog

Version 1.0.0

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

Description representing and modeling data in the NHGRI GWAS catalog

Enhances SNPlocs.Hsapiens.dbSNP.20110815, pd.genomewidesnp.6

Depends R (>= 2.14.0), methods, IRanges, GenomicRanges, snpStats,graph

Imports Biostrings

Suggests DO.db

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

License Artistic-2.0

LazyLoad yes

R topics documented:

`gwascat-package` *representing and modeling data in the NHGRI GWAS catalog*

Description

representing and modeling data in the NHGRI GWAS catalog, using GRanges and allied infrastructure

Details

Package:	gwascat
Version:	0.0.3
Suggests:	
Depends:	R (>= 2.14.0), methods, IRanges, GenomicRanges
Imports:	
License:	Artistic-2.0
LazyLoad:	yes
Built:	R 2.15.0; ; 2012-02-10 21:08:32 UTC; unix

Index:

Upon attachment, a `GRanges-class` structure call `gwrngs` is formed which can be interrogated by position or through use of element metadata to learn about catalogued GWAS associations.

The data objects

'g17SM' 'gg17N' 'gw6.rs_17' 'low17' 'rules_6.0_1kg_17'

are described in vignettes.

Author(s)

VJ Carey <stvjc@channing.harvard.edu>

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

References

`http://www.genome.gov/gwastudies/`.

Partial support from the Computational Biology Group at Genentech, Inc.

Examples

```
## Not run:
 gwrngs

## End(Not run)
```

`gwaswloc-class` *Class* `"gwaswloc"`

Description

A container for GRanges instances representing information in the NHGRI GWAS catalog.

Objects from the Class

Objects can be created by calls of the form `new("gwaswloc", ...)`. Any GRanges instance can be supplied.

Slots

- `extractDate`**:** character set manually in .onAttach code to indicate date of retrieval of base table
- `seqnames`**:** Object of class `"Rle"` typically representing chromosome numbers of loci associated with specific traits
- `ranges`**:** Object of class `"IRanges"` genomic coordinates for locus
- `strand`**:** Object of class `"Rle"` identifier of chromosome strand
- `elementMetadata`**:** Object of class `"DataFrame"` general `DataFrame-class` instance providing attributes for the locus-trait association
- `seqinfo`**:** Object of class `"Seqinfo"`
- `metadata`**:** Object of class `"list"`

Extends

Class `"GRanges"`, directly. Class `"GenomicRanges"`, by class "GRanges", distance 2. Class `"Vector"`, by class "GRanges", distance 3. Class `"GenomicRangesORmissing"`, by class "GRanges", distance 3. Class `"GenomicRangesORGRangesList"`, by class "GRanges", distance 3. Class `"Annotated"`, by class "GRanges", distance 4.

Methods

- **[** `signature(x = "gwaswloc")`: a character argument to the bracket will be assumed to be a dbSNP identifier for a SNP locus, and records corresponding to this SNP are extracted; numeric indexes are supported as for `GRanges-class` instances.
- **getRsids** `signature(x = "gwaswloc")`: extract all dbSNP identifiers as a character vector
- **getTraits** `signature(x = "gwaswloc")`: extract all traits (NHGRI term 'Disease/Trait') as a character vector
- **subsetByChromosome** `signature(x = "gwaswloc")`: select records by chromosome, a vector of chromosomes may be supplied
- **subsetByTraits** `signature(x = "gwaswloc")`: select all records corresponding to a given vector of traits

Note

In gwascat package, the globally accessible `gwaswloc` instance gwrngs is created upon attachment.

Author(s)

VJ Carey <stvjc@channing.harvard.edu>

References

http://www.genome.gov/gwastudies/

Examples

```
showClass("gwaswloc")
```

`gwdf_2012_02_02` *internal data frame for NHGRI GWAS catalog*

Description

convenience container for imported table from NHGRI GWAS catalog

Usage

```
data(gwdf_2012_02_02) # or more recent elements available
```

Format

A data frame with 7309 observations on the following 34 variables.

`Date Added to Catalog` a character vector

`PUBMEDID` a character vector

`First Author` a character vector

`Date` a character vector

`Journal` a character vector

`Link` a character vector

`Study` a character vector

`Disease/Trait` a character vector

`Initial Sample Size` a character vector

`Replication Sample Size` a character vector

`Region` a character vector

`Chr_id` a character vector

`Chr_pos` a character vector

`Reported Gene(s)` a character vector

`Mapped_gene` a character vector

`Upstream_gene_id` a character vector

`Downstream_gene_id` a character vector

`Snp_gene_ids` a character vector

`Upstream_gene_distance` a character vector

`Downstream_gene_distance` a character vector

`Strongest SNP-Risk Allele` a character vector

`SNPs` a character vector

`Merged` a character vector

`Snp_id_current` a character vector

`Context` a character vector

`Intergenic` a character vector

`Risk Allele Frequency` a character vector

`p-Value` a character vector

`Pvalue_mlog` a character vector

`p-Value (text)` a character vector

`OR or beta` a character vector

`95% CI (text)` a character vector

`Platform..` a character vector

`CNV` a character vector

Note

The .onAttach function specifies which data frame is transformed to GRanges.

Source

`http://www.genome.gov/gwastudies`

Examples

```
## Not run:
data(gwdf_2012_02_02)
data(gwdf_2012_03_07)

## End(Not run)
```

`locon6` *location information for 10000 SNPs probed on Affy GW 6.0*

Description

location information for 10000 SNPs probed on Affy GW 6.0

Usage

```
data(locon6)
```

Format

A data frame with 10000 observations on the following 3 variables.

`dbsnp_rs_id` a character vector

`chrom` a character vector

`physical_pos` a numeric vector

Details

extracted from pd.genomewidesnp.6 v 1.4.0; for demonstration purposes

Examples

```
data(locon6)
str(locon6)
```

`obo2graphNEL` *convert a typical OBO text file to a graphNEL instance (using Term elements)*

Description

convert a typical OBO text file to a graphNEL instance (using Term elements)

Usage

```
obo2graphNEL(obo, kill = "\\[Typedef\\]")
```

Arguments

`obo`	string naming a file in OBO format
`kill`	entity types to be excluded from processing – probably this should be in a 'keep' form, but for now this works.

Details

Very rudimentary list and grep operations are used to retain sufficient information to map the DAG to a graphNEL, using formal term identifiers as node names and 'is-a' relationships as edges, and term names and other metadata are assigned to nodeData components.

Value

a graphNEL instance

Note

The OBO for Human Disease ontology is serialized as text with this package.

Author(s)

VJ Carey <stvjc@channing.harvard.edu>

References

For use with human disease ontology, `http://www.obofoundry.org/cgi-bin/detail.cgi?id=disease_ontology`

`riskyAlleleCount` *given a matrix of subjects x SNP calls, count number of risky alleles*

Description

given a matrix of subjects x SNP calls, count number of risky alleles for various conditions, relative to NHGRI GWAS catalog

Usage

```
riskyAlleleCount(callmat, matIsAB = TRUE, chr,
  gwwl = gwrngs, snpap = "SNPlocs.Hsapiens.dbSNP.20110815",
  gencode = c("A/A", "A/B", "B/B"))
```

Arguments

`callmat`	matrix with subjects as rows, SNPs as columns; entries can be generic A/A, A/B, B/B, or specific nucleotide calls
`matIsAB`	logical, FALSE if nucleotide codes are present, TRUE if generic call codes are present; in the latter case, gwascat:::ABmat2nuc will be run
`chr`	code for chromosome, should work with the SNP annotation getSNPlocs function, so likely "ch[nn]"
`gwwl`	an instance of `gwaswloc`
`snpap`	name of a Bioconductor SNPlocs.Hsapiens.dbSNP.* package
`gencode`	codes used for generic SNP call

Value

matrix with rows corresponding to subjects , columns corresponding to SNP

Examples

```
if (!exists("gwrngs")) gwascat:::.onAttach("a", "b")
data(gg17N) # translated from GGdata chr 17 calls using ABmat2nuc
h17 = riskyAlleleCount(gg17N, matIsAB=FALSE, chr="ch17")
h17[1:5,1:5]
table(as.numeric(h17))
```

`topTraits` *operations on GWAS catalog*

Description

operations on GWAS catalog

Usage

```
topTraits (gwwl, n=10, tag="Disease.Trait")

locs4trait(gwwl, trait, tag="Disease.Trait")

chklocs(chrtag="20", gwwl=gwrngs)
```

Arguments

`gwwl`	instance of `gwaswloc`
`n`	numeric, number of traits to report
`tag`	character, name of field to be used for trait enumeration
`trait`	character, trait to use for filtering
`chrtag`	character, chromosome identifier

Value

`topTraits` returns a character vector of most frequently occurring traits in the database

`locs4trait` returns a `gwaswloc` object with records defining associations to the specified trait

`chklocs` returns a logical that is TRUE when the asserted locations of SNP in the GWAS catalog agree with the locations given in the dbSNP package SNPlocs.Hsapiens.dbSNP.20110815

Author(s)

VJ Carey <stvjc@channing.harvard.edu>

Examples

```
if (!exists("gwrngs")) gwascat:::.onAttach("a", "b")
topTraits(gwrngs)
```

Index