

# Package ‘restfulSE’

April 12, 2018

**Title** Access matrix-like HDF5 server content or BigQuery content through a SummarizedExperiment interface

**Description** This package provides functions and classes to interface with remote data stores by operating on SummarizedExperiment-like objects.

**Version** 1.0.2

**Suggests** knitr, testthat, Rtsne, org.Mm.eg.db, org.Hs.eg.db, BiocStyle

**Imports** utils, stats, methods, S4Vectors, DelayedArray, reshape2, AnnotationDbi, DBI, GO.db, rhdf5client, dplyr (>= 0.7.1), magrittr, bigrquery, ExperimentHub, AnnotationHub

**Depends** R (>= 3.4), SummarizedExperiment

**License** Artistic-2.0

**LazyLoad** yes

**biocViews** Infrastructure, SingleCell, Transcriptomics, Sequencing, Coverage

**RoxygenNote** 6.0.1.9000

**Collate** localReplaceSlots.R RESTfulSE.R demos.R seByTumor.R isbCgc.R

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Vincent Carey [aut],  
Shweta Gopaulakrishnan [cre, aut]

**Maintainer** Shweta Gopaulakrishnan <reshg@channing.harvard.edu>

## R topics documented:

assay,BQSummarizedExperiment,missing-method . . . . .	2
assayNames,BQSummarizedExperiment-method . . . . .	2
assays,RESTfulSummarizedExperiment-method . . . . .	3
BQSummarizedExperiment-class . . . . .	4
cgcConn . . . . .	4
dim,RESTfulSummarizedExperiment-method . . . . .	5
goPatt . . . . .	5
gtexTiss . . . . .	6
isbCgcDatasets . . . . .	7
isbCgcTables . . . . .	7
RESTfulSummarizedExperiment . . . . .	8

RESTfulSummarizedExperiment-class . . . . .	9
se100k . . . . .	9
seByTumor . . . . .	10
seByTumor_2016 . . . . .	11
tasicCortex . . . . .	12

## Index 13

---

assay,BQSummarizedExperiment,missing-method  
*extract assay data*

---

### Description

extract assay data

### Usage

```
## S4 method for signature 'BQSummarizedExperiment,missing'
assay(x, i, ...)
```

### Arguments

x	BQSummarizedExperiment instance
i	index for retrieval, ignored at present
...	not used

### Value

matrix

### Note

Very experimental approach to retrieving numerical data given a SummarizedExperiment 'shell'. We need more checking of consistency between assay and clinical data before creating the shell. We use dcast to transform query result to a matrix, and some 'individuals' may have multiple contributions ... we use fun.aggregate = max and will see warnings until this is cleared up.

---

assayNames,BQSummarizedExperiment-method  
*Placeholder for assay name extractor for a BQSummarizedExperiment instance.*

---

### Description

Placeholder for assay name extractor for a BQSummarizedExperiment instance.

### Usage

```
## S4 method for signature 'BQSummarizedExperiment'
assayNames(x, ...)
```

**Arguments**

x                    instance of `BQSummarizedExperiment`  
...                  not used

**Value**

string indicating that assay is served by BigQuery, nameless

**Note**

This function supplies a placeholder for this early version of a `SummarizedExperiment` instance to BigQuery. At present there is only one assay supported; future work will help to reduce special coding for BigQuery back end.

---

assays, *RESTfulSummarizedExperiment-method*

*Assays access for RESTfulSummarizedExperiment*

---

**Description**

Assays access for `RESTfulSummarizedExperiment`

**Usage**

```
## S4 method for signature 'RESTfulSummarizedExperiment'  
assays(x, ..., withDimnames = TRUE)
```

**Arguments**

x                    instance of `RESTfulSummarizedExperiment`  
...                  not used  
withDimnames      logical defaults to TRUE

**Value**

at present a `SimpleList` is returned as a dummy placeholder

---

BQSummarizedExperiment-class

*Define a class to use BigQuery data through SummarizedExperiment interface*

---

### Description

Define a class to use BigQuery data through SummarizedExperiment interface

### Slots

rowQref a BigQueryConnection wrapped in tbl\_dbi that holds rowData for the SummarizedExperiment instance

colQref a BigQueryConnection wrapped in tbl\_dbi that holds colData for the SummarizedExperiment instance

rowkey character(1) name of a field in the table referenced by rowQref to be used as key for features

colkey character(1) name of a field in the table referenced by colQref to use as key for samples

assaytbl character(1) name to be used to select table providing assay content

### Note

This is an experimental structure to probe the concept that one can use a SummarizedExperiment object to interact with BigQuery data, particularly TCGA data. The slots rowQref and colQref are expected to be BigQuery connections which supply information on features and samples respectively, in a way that is consistent with the assay representation. See [seByTumor](#) for illustration.

---

cgcConn

*Simplify connection to a BigQuery dataset for the project "isb-cgc"*

---

### Description

Simplify connection to a BigQuery dataset for the project "isb-cgc"

### Usage

```
cgcConn(dataset = "TCGA_bioclin_v0", project = "isb-cgc",
         billing = Sys.getenv("CGC_BILLING"))
```

### Arguments

dataset character string with dataset name

project character string with project name

billing character(1) with billing code

### Value

instance of [BigQueryConnection-class](#)

**Note**

This function operates on a BigQuery project to select a dataset and return a connection. If the google billing code is assigned to environment variable CGC\_BILLING, that will be used to authenticate the user and collect charges. Alternately the billing code can be given as a parameter.

**Examples**

```

cgcConn
require(bigrquery)
# defaults concern new GDC-compliant format
if (nchar(Sys.getenv("CGC_BILLING"))>0) {
  clin = cgcConn()
  dbListTables(clin)
}

```

---

dim,RESTfulSummarizedExperiment-method

*Dimension access for RESTfulSummarizedExperiment*

---

**Description**

Dimension access for RESTfulSummarizedExperiment

**Usage**

```

## S4 method for signature 'RESTfulSummarizedExperiment'
dim(x)

```

**Arguments**

x                    instance of RESTfulSummarizedExperiment

**Value**

vector of nrows, ncols

---

goPatt

*Create a data.frame with ENSEMBL and SYMBOL identifiers associated with a GO TERM specified by a regular expression in termPattern*

---

**Description**

Create a data.frame with ENSEMBL and SYMBOL identifiers associated with a GO TERM specified by a regular expression in termPattern

**Usage**

```

goPatt(termPattern = "neurotro", targets = c("ENSEMBL", "SYMBOL"),
       organism = "Hs", inst = "eg")

```

**Arguments**

termPattern	a character string encoding a regular expression to be matched to keys of type TERM in GO.db
targets	columns to be returned from org.[organism].[inst].db
organism	two-letter code for organism in the OrgDb family of packages
inst	two- or three-letter code (e.g., eg for ENTREZ GENE or sgd for yeastgenome.org) identifying institute responsible for annotation

**Value**

data.frame

**Examples**

```
gp = goPatt()
dim(gp)
head(gp)
```

---

gtexTiss	<i>Convenience function for access to gene-level GTEx tissues, as quantified in recount</i>
----------	---

---

**Description**

Convenience function for access to gene-level GTEx tissues, as quantified in recount

**Usage**

```
gtexTiss(url = "http://h5s.channingremotedata.org:5000", tag = "tissues")
```

**Arguments**

url	ip address/host for HDF5 server
tag	name of hdf5 file on server

**Value**

RESTfulSummarizedExperiment instance

**Examples**

```
gtexTiss()
```

---

isbCgcDatasets	<i>Access the dataset names in the "isb-cgc" project</i>
----------------	--

---

**Description**

Access the dataset names in the "isb-cgc" project

**Usage**

```
isbCgcDatasets()
```

**Value**

character vector

**Examples**

```
isbCgcDatasets() # might be out of date ... can use list_datasets in bigquery
```

---

isbCgcTables	<i>List the tables in a selected dataset</i>
--------------	--

---

**Description**

List the tables in a selected dataset

**Usage**

```
isbCgcTables(dataset = "TCGA_hg19_data_v0",  
             billing = Sys.getenv("CGC_BILLING"))
```

**Arguments**

dataset	character string identifying a table in "isb-cgc"
billing	Google BigQuery billing code, which can be set in an environment variable CGC_BILLING

**Value**

character vector

**Examples**

```
# be sure that .cgcBilling is set  
code = Sys.getenv("CGC_BILLING")  
if (!(nchar(code)==0)) {  
  isbCgcTables()  
}
```

---

 RESTfulSummarizedExperiment

*Construct RESTfulSummarizedExperiment*


---

### Description

Construct RESTfulSummarizedExperiment  
hidden constructor

### Usage

```
RESTfulSummarizedExperiment(se, source)

.RESTfulSummarizedExperiment(se, source)

## S4 method for signature 'RESTfulSummarizedExperiment,missing'
assay(x, i, ...)
```

### Arguments

se	SummarizedExperiment instance, assay component can be empty SimpleList
source	instance of H5S_dataset
x	instance of RESTfulSummarizedExperiment
i	not used
...	not used

### Value

instance of RESTfulSummarizedExperiment  
matrix

### Note

RESTfulSummarizedExperiment contains a global dimnames list generated at creation. It is possible that standard operations on a SummarizedExperiment will engender dimnames components that differ from the initial global dimnames, principally through unification (adding suffixes when dimname elements are repeated). When this is detected, assay() will fail with a complaint about length(setdiff(\*names(x), x@globalDimnames[[...]])).

### Examples

```
require("rhdf5client")
bigec2 = H5S_source(serverURL="http://h5s.channingremotedata.org:5000")
banoh5 = bigec2[["assays"]] # banovichSE
ehub = ExperimentHub::ExperimentHub()
myfiles <- AnnotationHub::query(ehub , "restfulSEData")
myfiles[["EH551"]] -> banoSEMeta
rr = RESTfulSummarizedExperiment(banoSEMeta, banoh5)
rr
rr2 = rr[1:4, 1:5] # just modify metadata
rr2
assay(rr2) # extract data
```



---

 RESTfulSummarizedExperiment-class

*HDF5Server-based assay for SummarizedExperiment*


---

**Description**

HDF5Server-based assay for SummarizedExperiment

**Usage**

```
## S4 method for signature 'RESTfulSummarizedExperiment'
assayNames(x, ...)
```

```
## S4 method for signature 'RESTfulSummarizedExperiment,numeric,numeric,ANY'
x[i, j, ...,
  drop = FALSE]
```

**Arguments**

x	instance of RESTfulSummarizedExperiment
...	not used
i	numeric selection vector
j	numeric selection vector
drop	not used

**Value**

instance of RESTfulSummarizedExperiment

---

se100k	<i>Convenience functions using EC2 server to extract tenx neurons full or subset data</i>
--------	---

---

**Description**

Convenience functions using EC2 server to extract tenx neurons full or subset data

**Usage**

```
se100k(url = "http://h5s.channingremotedata.org:5000",
  tag = "tenx_100k_sorted")
```

```
se1.3M(url = "http://h5s.channingremotedata.org:5000", tag = "tenx_full")
```

**Arguments**

url	server URL
tag	string giving the internal dataset name

**Value**

RESTfulSummarizedExperiment

**Note**

se1.3M provides access to the full 1.3 million neurons with features in their order as given in the original HDF5 while se100k provides access to only 100k neurons with expression features sorted by genomic location

**Examples**

```
ss = se100k()
# get a set of genes from Tasic et al. 2016 Nature Neuroscience
tc = tasicCortex()
adultCort = tc$GENEID
# subset
csums = apply(assay(ss[adultCort,1:500]),1,sum)
names(csums) = tc$SYMBOL
csums
```

---

seByTumor

*Given a BigQueryConnection to the 2017 GDC-oriented ISB TCGA bigtables, obtain a SummarizedExperiment 'shell' rowData and colData*

---

**Description**

Given a BigQueryConnection to the 2017 GDC-oriented ISB TCGA bigtables, obtain a SummarizedExperiment 'shell' rowData and colData

**Usage**

```
seByTumor(tumorCode = "LUAD", assayTblName = "RNAseq_Gene_Expression",
  rdColsToKeep = c("gene_name", "Ensembl_gene_id", "gene_type"),
  bqConnClinical, bqConnAssay, rowkey = "Ensembl_gene_id",
  colkey = "case_barcode", assayvbl = "HTSeq__Counts")
```

**Arguments**

tumorCode	one of the concise TCGA codes in a character string – not checked, defaults to "LUAD", lung adenocarcinoma
assayTblName	the name of the assay whose annotation will be used as rowData
rdColsToKeep	columns of assay table to use in rowData component
bqConnClinical	instance of BigQueryConnection from bigrquery, for access to clinical metadata – current expectation is that the BigQuery dataset is named "TCGA_bioclin_v0" and has a table called "Clinical"
bqConnAssay	instance of BigQueryConnection from bigrquery – current expectation is that the BigQuery dataset is named "TCGA_hg19_data_v0"
rowkey	name of a field to be used as key for rows
colkey	name of a field to use as key for samples
assayvbl	name of field to use for numerical values

**Value**

SummarizedExperiment

**Note**

This function demonstrates the use of external resources for rowData, colData and assay components of a SummarizedExperiment instance. The intention is that the full complement of activities supported by [SummarizedExperiment-class](#) are likewise supported through this class, with assay data and sample and feature metadata all external and in BigQuery projects. The seByTumor function is provided to generate an example of this approach with minimal user configuration.

**Examples**

```
require(bigrquery)
# be sure that .cgcBilling is set
code = Sys.getenv("CGC_BILLING")
if (!(nchar(code)==0)) {
  clinQ = cgcConn(billing=code)
  assayQ = cgcConn( dataset = "TCGA_hg38_data_v0", billing=code )
  myexpShell = seByTumor( bqConnClinical=clinQ,
                        bqConnAssay=assayQ)
  print(myexpShell)
  print(nrow(myexpShell) == 60483)
  print(ncol(myexpShell) == 522)
  assay(myexpShell[11:15,1:4]) # some case_barcodes repeat
}
```

seByTumor\_2016

*Given a BigQueryConnection to the 2016 ISB TCGA bigtables, obtain a SummarizedExperiment 'shell' rowData and colData*

**Description**

Given a BigQueryConnection to the 2016 ISB TCGA bigtables, obtain a SummarizedExperiment 'shell' rowData and colData

**Usage**

```
seByTumor_2016(tumorCode = "LUAD", assayTblName = "mRNA_UNC_HiSeq_RSEM",
               rdColsToKeep = c("original_gene_symbol", "HGNC_gene_symbol", "gene_id",
                                "Study"), bqConn)
```

**Arguments**

tumorCode	one of the concise TCGA codes in a character string – not checked, defaults to "LUAD", lung adenocarcinoma
assayTblName	the name of the assay whose annotation will be used as rowData
rdColsToKeep	columns of assay table to use in rowData component
bqConn	instance of BigQueryConnection from bigrquery

**Value**

SummarizedExperiment instance, with BigQuery reference as assay

---

tasicCortex	<i>A set of mouse cortex marker genes.</i>
-------------	--

---

**Description**

A set of mouse cortex marker genes.

**Usage**

```
tasicCortex()
```

**Value**

data.frame with columns SYMBOL, GENEID

**Note**

<http://www.nature.com/doi/finder/10.1038/nn.4216>, Fig 1C

**Examples**

```
head(tasicCortex())
```

# Index

.RESTfulSummarizedExperiment  
(RESTfulSummarizedExperiment),  
8

[,RESTfulSummarizedExperiment,numeric,numeric,ANY-method  
(RESTfulSummarizedExperiment-class),  
9

assay (RESTfulSummarizedExperiment), 8

assay,BQSummarizedExperiment,missing-method,  
2

assay,RESTfulSummarizedExperiment,missing-method  
(RESTfulSummarizedExperiment),  
8

assayNames,BQSummarizedExperiment-method,  
2

assayNames,RESTfulSummarizedExperiment-method  
(RESTfulSummarizedExperiment-class),  
9

assays,RESTfulSummarizedExperiment-method,  
3

BQSummarizedExperiment-class, 4

cgcConn, 4

dim,RESTfulSummarizedExperiment-method,  
5

goPatt, 5

gtexTiss, 6

isbCgcDatasets, 7

isbCgcTables, 7

RESTfulSummarizedExperiment, 8

RESTfulSummarizedExperiment,RangedSummarizedExperiment,H5S\_dataset-method  
(RESTfulSummarizedExperiment),  
8

RESTfulSummarizedExperiment-class, 9

se1.3M (se100k), 9

se100k, 9

seByTumor, 4, 10

seByTumor\_2016, 11

tasieCortex, 12