

restfulSEData – SummarizedExperiment shells for remote assay data

Vincent J. Carey, stvjc at channing.harvard.edu, Shweta Gopaulakrishnan reshg at channing.harvard.edu

November 05, 2019

Contents

1	Introduction	2
---	------------------------	---

1 Introduction

This package includes SummarizedExperiment or RangedSummarizedExperiment instances from which assay data has been removed, so that it can be restored in real time from remote stores.

- The example given below elaborates on how to access the data from the standard ExperimentHub interface :

```
library(ExperimentHub)
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:parallel':
##
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB
## The following objects are masked from 'package:stats':
##
##   IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
##   Filter, Find, Map, Position, Reduce, anyDuplicated, append,
##   as.data.frame, basename, cbind, colnames, dirname, do.call,
##   duplicated, eval, evalq, get, grep, grepl, intersect,
##   is.unsorted, lapply, mapply, match, mget, order, paste, pmax,
##   pmax.int, pmin, pmin.int, rank, rbind, rownames, sapply,
##   setdiff, sort, table, tapply, union, unique, unsplit, which,
##   which.max, which.min
## Loading required package: AnnotationHub
## Loading required package: BiocFileCache
## Loading required package: dbplyr
ehub = ExperimentHub()
## snapshotDate(): 2019-10-22
myfiles <- query(ehub , "restfulSEData")
myfiles
## ExperimentHub with 7 records
## # snapshotDate(): 2019-10-22
## # $dataprovder: 10x Genomics, GEO, GTex, Illumina 450 methylation assay,...
## # $species: Homo sapiens, Mus musculus (E18 mice), Mus musculus
## # $rdataclass: RangedSummarizedExperiment, DataFrame, GRanges
## # additional mcols(): taxonomyid, genome, description,
## #   coordinate_1_based, maintainer, rdatadateadded, preparerclass,
## #   tags, rdatapath, sourceurl, sourcetype
## # retrieve records with, e.g., 'object[["EH551"]]'
##
##           title
## EH551 | banoSEMeta
## EH552 | st100k
```

restfulSEData – SummarizedExperiment shells for remote assay data

```
## EH553 | st400k
## EH555 | gr450k
## EH556 | gtexRecount
## EH557 | tasicST6
## EH1656 | full_1Mneurons
myfiles[[1]]
## see ?restfulSEData and browseVignettes('restfulSEData') for documentation
## loading from cache
## class: RangedSummarizedExperiment
## dim: 329469 64
## metadata(0):
## assays(1): betas
## rownames(329469): cg00000029 cg00000165 ... ch.9.98989607R
## ch.9.991104F
## rowData names(10): addressA addressB ... probeEnd probeTarget
## colnames(64): NA18498 NA18499 ... NA18489 NA18909
## colData names(35): title geo_accession ... data_row_count naid
myfiles[["EH551"]] #load by EH id
## see ?restfulSEData and browseVignettes('restfulSEData') for documentation
## loading from cache
## class: RangedSummarizedExperiment
## dim: 329469 64
## metadata(0):
## assays(1): betas
## rownames(329469): cg00000029 cg00000165 ... ch.9.98989607R
## ch.9.991104F
## rowData names(10): addressA addressB ... probeEnd probeTarget
## colnames(64): NA18498 NA18499 ... NA18489 NA18909
## colData names(35): title geo_accession ... data_row_count naid
```

- To get a flavor of the datasets present :

```
dataResource()
## [1] banoSEMata : Metadata RangedSummarizedExperiment shell for banovichSE
## [2] st100k : Metadata RangedSummarizedExperiment shell for 100k cells from 10x genomics 1.3 million neuron
## [3] st400k : Metadata RangedSummarizedExperiment shell for 400k cells range-sorted from 10xgenomics 1.3 m
## [4] full_1Mneurons : Metadata RangedSummarizedExperiment shell for the full 1.3 million neuron dataset fro
## [5] gr450k : GRanges with metadata for illumina 450k methylation assay
## [6] gtexRecount : Metadata RangedSummarizedExperiment shell for RECOUNT gtex rse_gene
## [7] tasicST6 : Supplemental table from Tasic et al. 2016
## 7 Levels: banoSEMata : Metadata RangedSummarizedExperiment shell for banovichSE ...
```

- To use the data with “restfulSE” R package : We grab the ExperimentHub ID of the dataset we are interested in.

```
myfiles[["EH551"]] -> banoSEMata
banoSEMata
```