restfulSE – experiments with HDF5 server content wrapped in SummarizedExperiment

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

April 27, 2017

Contents

1   Introduction

Extensive human and computational effort is expended on downloading and managing large genomic data at site of analysis. Interoperable formats that are accessible via generic operations like those in RESTful APIs may help to improve cost-effectiveness of genome-scale analyses.

In this report we examine the use of HDF5 server as a back end for assay data, mediated through the RangedSummarizedExperiment API for interactive use.

A modest server configured to deliver HDF5 content via a RESTful API has been prepared and is used in this vignette.

2   Background

Banovich et al. published a subset of DNA methylation measures assembled on 64 samples of immortalized B-cells from the YRI HapMap cohort.

```
library(restfulSE)
data(banoSEMeta)
banoSEMeta
## class: RangedSummarizedExperiment
## dim: 329469 64
## metadata(0):
## assays(0):
## 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
```

The numerical data have been exported using H. Pages' saveHDF5SummarizedExperiment applied to the banovichSE SummarizedExperiment in the yriMulti package. The HDF5 component is simply copied into the server data space on the remote server. A wrapper class has been defined in the restfulSE package.

```
banoh5   = banoH5("http://54.163.220.201:5000")
banoh5
## An HDF5-served dataset with internal dimensions 64 x 329469
## Northwest 5 x 5 submatrix:
##              [,1]      [,2]       [,3]       [,4]        [,5]
## [1,]   0.4733963 1.2943041 -0.8084735  1.0056166 -0.90349832
## [2,]   1.2364086 0.2099817 -0.2683763 -1.2354204 -0.02221144
## [3,] -0.2225818 1.6236857 -0.8654838 -0.3172301 -0.02754413
## [4,]   0.6572058 0.5527470 -1.8458295  0.6250719  1.17781474
## [5,] -0.1506308 0.7498020  0.3254333  1.3778200  0.59669990
```

3   Construction

```
restBano = RESTfulSummarizedExperiment(banoSEMeta, banoh5)
restBano
## class: RESTfulSummarizedExperiment
## dim: 329469 64
## metadata(0):
## assays(1): (served by HDF5Server)
## 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
```

4   Subsetting and assay extraction

Targeted extraction is possible, but index processing needs considerable work.

```
subr = restBano[1:4,1:6]
subr
## class: RESTfulSummarizedExperiment
## dim: 329469 6
## metadata(0):
## assays(1): (served by HDF5Server)
## rownames(4): cg00000029 cg00000165 cg00000236 cg00000289
## rowData names(10): addressA addressB ... probeEnd probeTarget
## colnames(6): NA18498 NA18499 ... NA18516 NA18517
## colData names(35): title geo_accession ... data_row_count naid
assay(subr)
##                 NA18498     NA18499     NA18501     NA18502     NA18516
## cg00000029   0.4733963   1.2364086 -0.2225818   0.6572058 -0.1506308
## cg00000165   1.2943041   0.2099817  1.6236857   0.5527470  0.7498020
## cg00000236 -0.8084735 -0.2683763 -0.8654838 -1.8458295  0.3254333
## cg00000289   1.0056166 -1.2354204 -0.3172301   0.6250719  1.3778200
##                 NA18517
## cg00000029 -0.189241860
## cg00000165 -1.515388273
## cg00000236  0.003436888
```

```
## cg00000289 -0.668499289
```