

# h5SE – experiments with delayed HDF5 export/import for SummarizedExperiments

Vincent J. Carey, *stvjc at channing.harvard.edu*

February 16, 2017

## Contents

---

1	Introduction	1
2	Export a standard RangedSummarizedExperiment	1
3	Import a packaged HDF5RangedSummarizedExperiment	2

## 1 Introduction

---

We want to get some background on the use of HDF5 and DelayedArray infrastructure for modest-size assay sets.

We'll work with some 450k data.

```
suppressPackageStartupMessages({
  library(SummarizedExperiment)
  library(HDF5Array)
  library(yriMulti)
  if (!exists("banovichSE")) data(banovichSE)
})
banovichSE
## class: RangedSummarizedExperiment
## dim: 329469 64
## metadata(0):
## assays(1): betas
## rownames(329469): cg000000029 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
```

## 2 Export a standard RangedSummarizedExperiment

---

Let's use the newly defined method for exporting the assay component of a SummarizedExperiment:

```
library(bench00M)
tf = tempfile()
dir.create(tf)
banoSE = saveHDF5SummarizedExperiment(banovichSE, tf, replace=TRUE)
banoSE
## class: RangedSummarizedExperiment
## dim: 329469 64
```

```
## metadata(0):
## assays(1): betas
## rownames(329469): cg000000029 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
```

Access capabilities:

```
assay(banoSE[1,])
## DelayedMatrix object of 1 x 64 doubles:
##      NA18498   NA18499   NA18501   .   NA18489   NA18909
## cg000000029  0.4733963  1.2943041 -0.8084735   .   0.6708168 -0.8609302
subsetByOverlaps(banoSE, rowRanges(banoSE)[1001:1010])
## class: RangedSummarizedExperiment
## dim: 10 64
## metadata(0):
## assays(1): betas
## rownames(10): cg00063006 cg00063040 ... cg00063346 cg00063357
## rowData names(10): addressA addressB ... probeEnd probeTarget
## colnames(64): NA18498 NA18499 ... NA18489 NA18909
## colData names(35): title geo_accession ... data_row_count naid
```

### 3 Import a packaged HDF5RangedSummarizedExperiment

---

```
litg = getLitGeu()
litg
## class: RangedSummarizedExperiment
## dim: 100 462
## metadata(3): MIAME constrHist colDataSource
## assays(1): exprs
## rownames(100): ENSG00000152931.6 ENSG00000183696.9 ...
##   ENSG00000239872.1 ENSG00000179023.6
## rowData names(18): source type ... tag ccidsid
## colnames(462): HG00096 HG00097 ... NA20826 NA20828
## colData names(35): Source.Name Comment.ENA_SAMPLE. ...
##   Factor.Value.laboratory. popcode
```