Bioconductor meetup November 20th: Creating a Bioconductor Package - Part 1.1

Shweta Gopaulakrishnan

11/13/2018

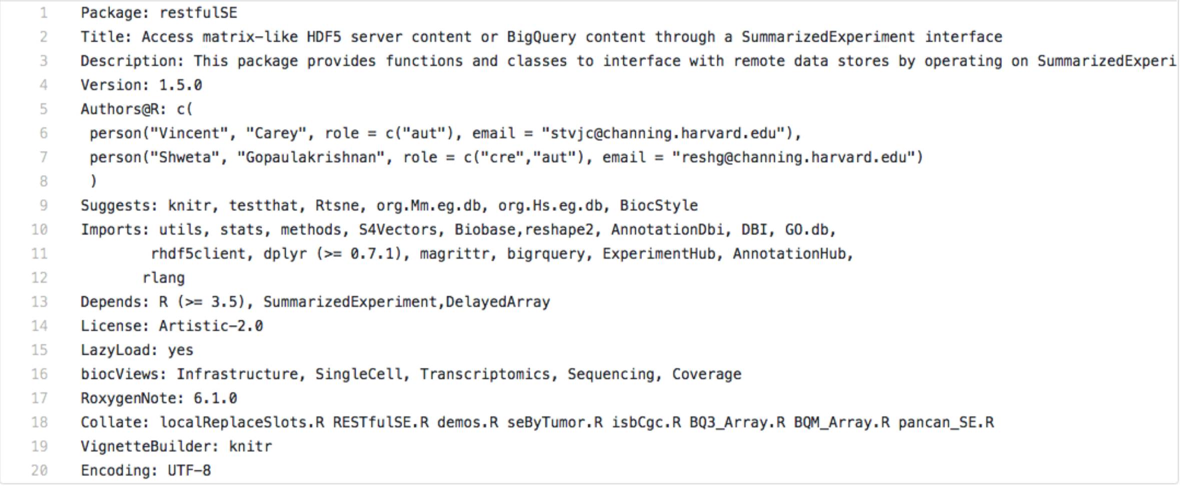
Code organization

Skeleton

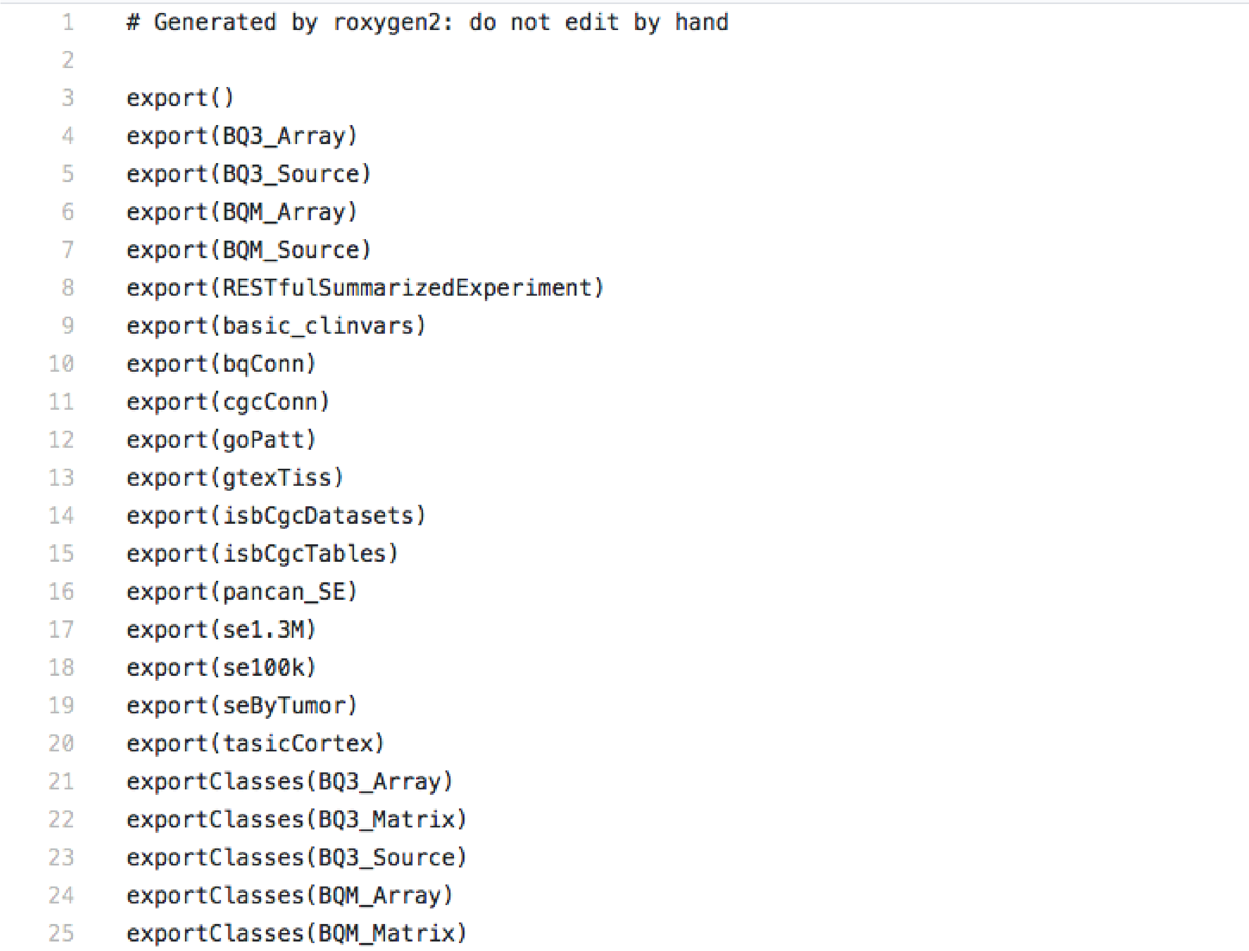
- DESCRIPTION : To store important metadata about your package.
- NAMESPACE : Defines the functions, classes, and methods that are imported into the name space, and exported for users.
- NEWS : Keeps track of changes to the code from one version to the next.
- CITATION : Appropriate citations must be included in help pages and vignettes; this aspect of documentation is no different from any scientific endeavor.
- Including Data
- Package Documentation (vignettes, man pages)
- Unit Tests
- .gitignore : Bioconductor requires a git repository for submission.

DESCRIPTION

- The job of the DESCRIPTION file is to store important metadata about your package



```
Package: restfulSE
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 SummarizedExperi
Version: 1.5.0
Authors@R: c(
 person("Vincent", "Carey", role = c("aut"), email = "stvjc@channing.harvard.edu"),
 person("Shweta", "Gopaulakrishnan", role = c("cre","aut"), email = "reshg@channing.harvard.edu")
 )
Suggests: knitr, testthat, Rtsne, org.Mm.eg.db, org.Hs.eg.db, BiocStyle
Imports: utils, stats, methods, S4Vectors, Biobase,reshape2, AnnotationDbi, DBI, GO.db,
        rhdf5client, dplyr (>= 0.7.1), magrittr, bigrquery, ExperimentHub, AnnotationHub,
       rlang
Depends: R (>= 3.5), SummarizedExperiment,DelayedArray
License: Artistic-2.0
LazyLoad: yes
biocViews: Infrastructure, SingleCell, Transcriptomics, Sequencing, Coverage
RoxygenNote: 6.1.0
Collate: localReplaceSlots.R RESTfulSE.R demos.R seByTumor.R isbCgc.R BQ3_Array.R BQM_Array.R pancan_SE.R
VignetteBuilder: knitr
Encoding: UTF-8
```



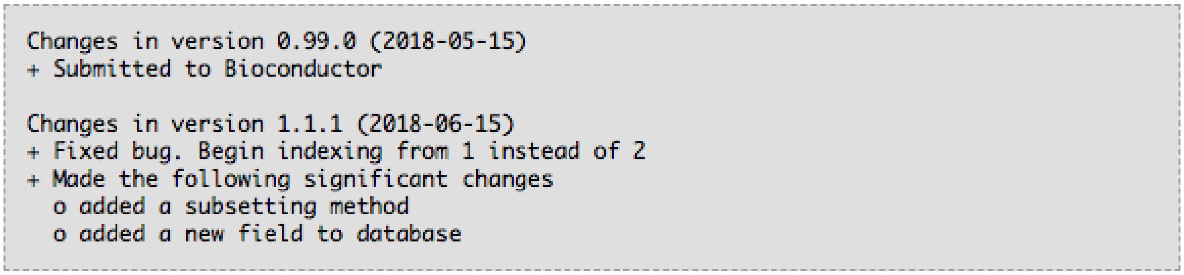
```
# Generated by roxygen2: do not edit by hand

export()
export(BQ3_Array)
export(BQ3_Source)
export(BQM_Array)
export(BQM_Source)
export(RESTfulSummarizedExperiment)
export(basic_clinvars)
export(bqConn)
export(cgcConn)
export(goPatt)
export(gtexTiss)
export(isbCgcDatasets)
export(isbCgcTables)
export(pancan_SE)
export(se1.3M)
export(se100k)
export(seByTumor)
export(tasicCortex)
exportClasses(BQ3_Array)
exportClasses(BQ3_Matrix)
exportClasses(BQ3_Source)
exportClasses(BQM_Array)
exportClasses(BQM_Matrix)
```

Figure 1:

NAMESPACE

NEWS

- A NEWS file should be included to keep track of changes to the code from one version to the next.



```
Changes in version 0.99.0 (2018-05-15)
+ Submitted to Bioconductor

Changes in version 1.1.1 (2018-06-15)
+ Fixed bug. Begin indexing from 1 instead of 2
+ Made the following significant changes
  o added a subsetting method
  o added a new field to database
```

Roxygen

Why do we need documentation?

Documentation is one of the most important aspects of a good package. Without it, users won't know how to use your package. Documentation is also useful for future-you (so you remember what your functions were supposed to do), and for developers extending your package.

How to use Roxygen ?

- Add roxygen comments to your .R files.
- Run devtools::document() (or roxygen2::roxygenise()) to convert roxygen comments to .Rd files. (devtools::document() calls roxygen2::roxygenise() to do the hard work.)
- Roxygen will also create a NAMESPACE for you, and will manage the Collate field in DESCRIPTION.

Documenting Functions

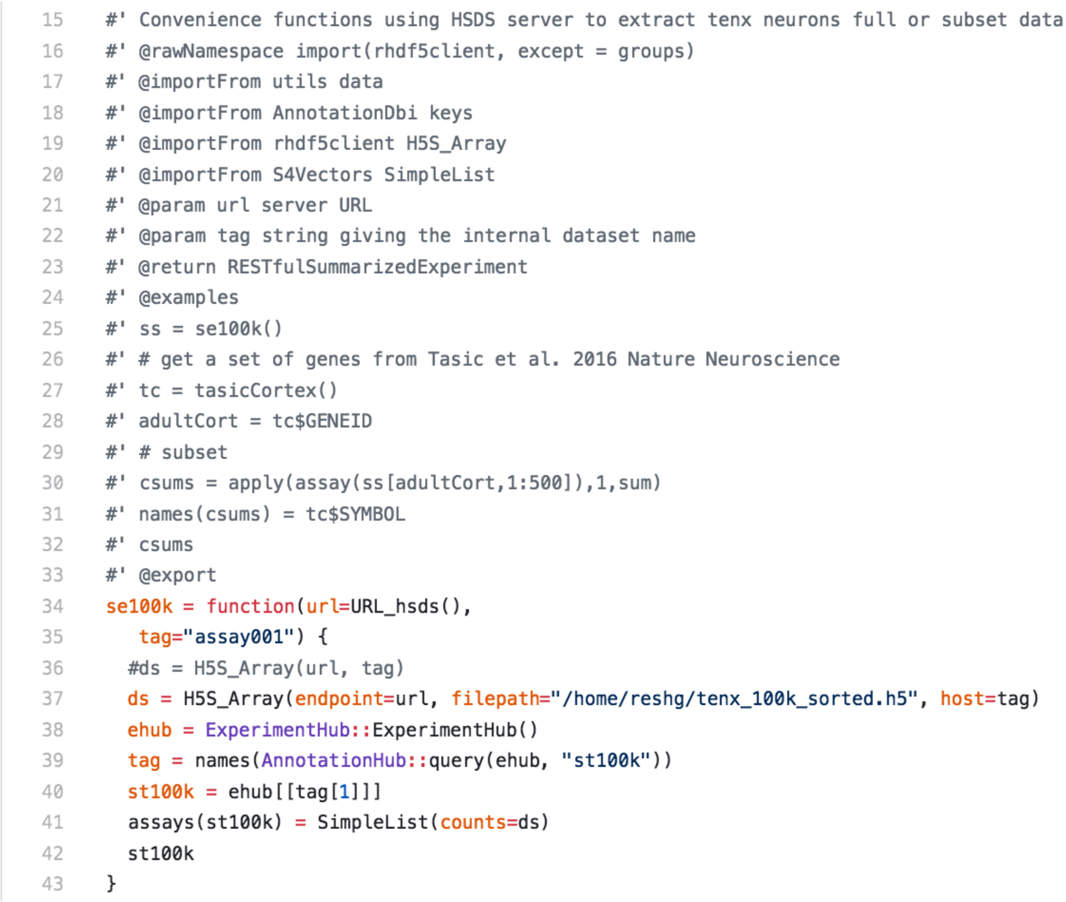
Documenting datasets

Documenting classes, generics and methods

Unit Tests

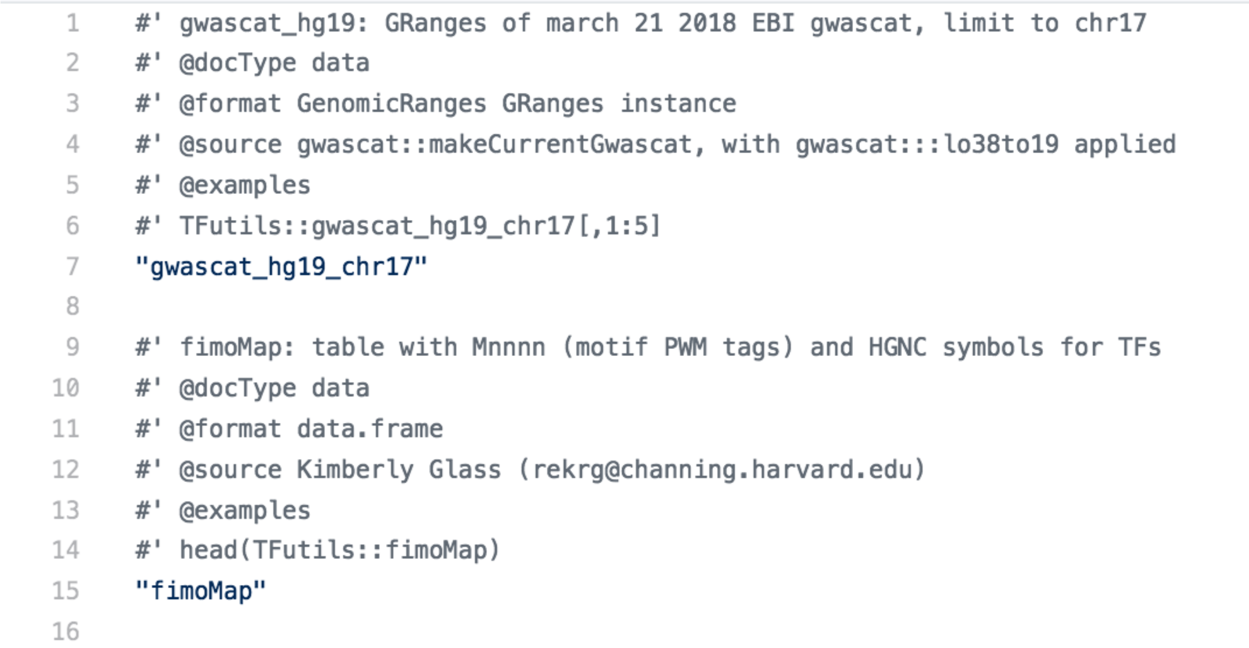
- Testing is a vital part of package development.
- It ensures that your code does what you want it to do.
- RUnit or testthat packages to write unit tests.

Guide : http://bioconductor.org/developers/how-to/unitTesting-guidelines/



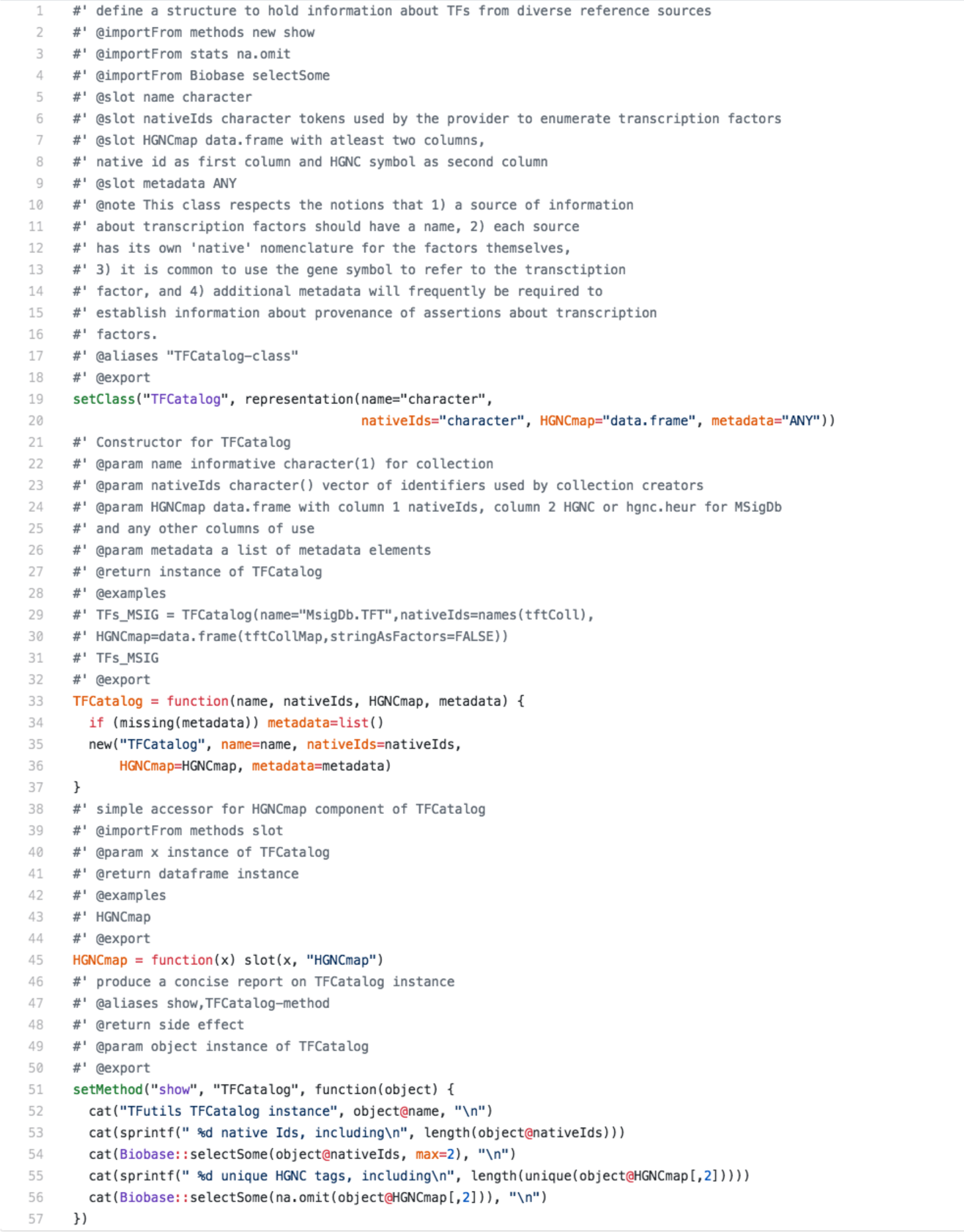
```r
#' Convenience functions using HSDS server to extract tenx neurons full or subset data
#' @rawNamespace import(rhdf5client, except = groups)
#' @importFrom utils data
#' @importFrom AnnotationDbi keys
#' @importFrom rhdf5client H5S_Array
#' @importFrom S4Vectors SimpleList
#' @param url server URL
#' @param tag string giving the internal dataset name
#' @return RESTfulSummarizedExperiment
#' @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
#' @export
se100k = function(url=URL_hsds(),
   tag="assay001") {
  #ds = H5S_Array(url, tag)
  ds = H5S_Array(endpoint=url, filepath="/home/reshg/tenx_100k_sorted.h5", host=tag)
  ehub = ExperimentHub::ExperimentHub()
  tag = names(AnnotationHub::query(ehub, "st100k"))
  st100k = ehub[[tag[1]]]
  assays(st100k) = SimpleList(counts=ds)
  st100k
}
```

Figure 2:



```r
#' gwascat_hg19: GRanges of march 21 2018 EBI gwascat, limit to chr17
#' @docType data
#' @format GenomicRanges GRanges instance
#' @source gwascat::makeCurrentGwascat, with gwascat:::lo38to19 applied
#' @examples
#' TFutils::gwascat_hg19_chr17[,1:5]
"gwascat_hg19_chr17"

#' fimoMap: table with Mnnnn (motif PWM tags) and HGNC symbols for TFs
#' @docType data
#' @format data.frame
#' @source Kimberly Glass (rekrg@channing.harvard.edu)
#' @examples
#' head(TFutils::fimoMap)
"fimoMap"

```

Figure 3:



```
#' define a structure to hold information about TFs from diverse reference sources
#' @importFrom methods new show
#' @importFrom stats na.omit
#' @importFrom Biobase selectSome
#' @slot name character
#' @slot nativeIds character tokens used by the provider to enumerate transcription factors
#' @slot HGNCmap data.frame with atleast two columns,
#' native id as first column and HGNC symbol as second column
#' @slot metadata ANY
#' @note This class respects the notions that 1) a source of information
#' about transcription factors should have a name, 2) each source
#' has its own 'native' nomenclature for the factors themselves,
#' 3) it is common to use the gene symbol to refer to the transctiption
#' factor, and 4) additional metadata will frequently be required to
#' establish information about provenance of assertions about transcription
#' factors.
#' @aliases "TFCatalog-class"
#' @export
setClass("TFCatalog", representation(name="character",
                                      nativeIds="character", HGNCmap="data.frame", metadata="ANY"))
#' Constructor for TFCatalog
#' @param name informative character(1) for collection
#' @param nativeIds character() vector of identifiers used by collection creators
#' @param HGNCmap data.frame with column 1 nativeIds, column 2 HGNC or hgnc.heur for MSigDb
#' and any other columns of use
#' @param metadata a list of metadata elements
#' @return instance of TFCatalog
#' @examples
#' TFs_MSIG = TFCatalog(name="MsigDb.TFT",nativeIds=names(tftColl),
#' HGNCmap=data.frame(tftCollMap,stringAsFactors=FALSE))
#' TFs_MSIG
#' @export
TFCatalog = function(name, nativeIds, HGNCmap, metadata) {
  if (missing(metadata)) metadata=list()
  new("TFCatalog", name=name, nativeIds=nativeIds,
      HGNCmap=HGNCmap, metadata=metadata)
}
#' simple accessor for HGNCmap component of TFCatalog
#' @importFrom methods slot
#' @param x instance of TFCatalog
#' @return dataframe instance
#' @examples
#' HGNCmap
#' @export
HGNCmap = function(x) slot(x, "HGNCmap")
#' produce a concise report on TFCatalog instance
#' @aliases show,TFCatalog-method
#' @return side effect
#' @param object instance of TFCatalog
#' @export
setMethod("show", "TFCatalog", function(object) {
  cat("TFutils TFCatalog instance", object@name, "\n")
  cat(sprintf(" %d native Ids, including\n", length(object@nativeIds)))
  cat(Biobase::selectSome(object@nativeIds, max=2), "\n")
  cat(sprintf(" %d unique HGNC tags, including\n", length(unique(object@HGNCmap[,2]))))
  cat(Biobase::selectSome(na.omit(object@HGNCmap[,2])), "\n")
})
```

Figure 4:

Test Coverage

- Test coverage refers to the percentage of your package code that is tested by your unit tests.
- Packages with higher coverage have a lower chance of containing bugs.

References

1. https://cran.r-project.org/web/packages/qwraps2/vignettes/create_pkg.html
2. http://r-pkgs.had.co.nz/
3. https://www.bioconductor.org/developers/package-guidelines/#unittest