Package ‘cellbaseR’

June 4, 2026

Type Package

Date 2016-04-17

Title Querying annotation data from the high performance Cellbase web

Version 1.37.0

Author Mohammed OE Abdallah

Maintainer Mohammed OE Abdallah `<melsiddieg@gmail.com>`

URL `https://github.com/melsiddieg/cellbaseR`

Description This R package makes use of the exhaustive RESTful Web service API that has been implemented for the Cellabase database. It enable researchers to query and obtain a wealth of biological information from a single database saving a lot of time. Another benefit is that researchers can easily make queries about different biological topics and link all this information together as all information is integrated.

License Apache License (== 2.0)

Depends R(>= 3.4)

Imports methods, jsonlite, httr, data.table, pbapply, tidyr, R.utils, Rsamtools, BiocParallel, foreach, utils, parallel, doParallel

Suggests BiocStyle, knitr, rmarkdown, Gviz, VariantAnnotation

RoxygenNote 7.2.3

biocViews Annotation, VariantAnnotation

Lazy TRUE

LazyData TRUE

Encoding UTF-8

Collate 'commons.R' 'AllClasses.R' 'AllGenerics.R' 'AnnotateVcf-methods.R' 'CellBaseParam-methods.R' 'CellBaseR-methods.R' 'cellbase.R' 'getCellbase-methods.R' 'getChromosomeInfo-methods.R' 'getClinical-methods.R' 'getGene-methods.R' 'getMeta-methods.R' 'getProtein-methods.R' 'getRegion-methods.R' 'getTranscript-methods.R' 'getVariant-methods.R' 'getXref-methods.R' 'show-methods.R' 'tools.R' 'user.R'

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/cellbaseR

git_branch devel

git_last_commit 01b61c7

git_last_commit_date 2026-04-28

Repository Bioconductor 3.24

Date/Publication 2026-06-04

Contents

`cellbaseR-package` *cellbaseR*

Description

Querying annotation data from the high performance Cellbase web services

Details

Documentation for the cellbaseR package

This R package makes use of the exhaustive RESTful Web service API that has been implemented for the Cellabase database. It enables researchers to query and obtain a wealth of biological information from a single database saving a lot of time. Another benefit is that researchers can easily make queries about different biological topics and link all this information together as all information is integrated. Currently Homo sapiens, Mus musculus and other 20 species are available and many others will be included soon. Results returned from the cellbase queries are parsed into R data.frames and other common R data strctures so users can readily get into downstream anaysis.

Author(s)

Mohammed OE Abdallah

See Also

Useful links:

- `https://github.com/melsiddieg/cellbaseR`

`AnnotateVcf,CellBaseR-method` *AnnotateVcf*

Description

This method is a convience method to annotate bgzipped tabix-indexed vcf files. It should be ideal for annotating small to medium sized vcf files.

Usage

```
## S4 method for signature 'CellBaseR'
AnnotateVcf(object, file, batch_size, num_threads, BPPARAM = bpparam())
```

Arguments

`object`	an object of class CellBaseR
`file`	Path to a bgzipped and tabix indexed vcf file
`batch_size`	intger if multiple queries are raised by a single method call, e.g. getting annotation info for several genes, queries will be sent to the server in batches. This slot indicates the size of each batch, e.g. 200
`num_threads`	number of asynchronus batches to be sent to the server
`BPPARAM`	a BiocParallel class object

Value

a dataframe with the results of the query

See Also

`https://github.com/opencb/cellbase/wiki` and the RESTful API documentation `http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/`

Examples

```
cb <- CellBaseR()
fl <- system.file("extdata", "hapmap_exome_chr22_200.vcf.gz",
                  package = "cellbaseR" )
res <- AnnotateVcf(object=cb, file=fl, BPPARAM = bpparam(workers=2),batch_size=100)
```

`CellBaseParam` *A Constructor for the CellBaseParam Object*

Description

CellBaseParam object is used to control what results are returned from the CellBaseR methods

Usage

```
CellBaseParam(
  assembly = character(),
  feature = character(),
  region = character(),
  rsid = character(),
  accession = character(),
  type = character(),
  mode_inheritance_labels = character(),
  clinsig_labels = character(),
  alleleOrigin = character(),
  consistency_labels = character(),
  so = character(),
  source = character(),
  trait = character(),
  include = character(),
  exclude = character(),
  limit = character()
)
```

Arguments

`assembly`	A character the assembly build to query, e.g.GRCh37(default)
`feature`	A character vector denoting the feature/s to be queried
`region`	A character vector denoting the region/s to be queried must be in the form 1:100000-1500000
`rsid`	A character vector denoting the rs ids to be queried
`accession`	A caharcter vector of Cinvar accessions
`type`	A caharcter vector of Variant types
`mode_inheritance_labels`	A character vector
`clinsig_labels`	A character vector
`alleleOrigin`	A character vector
`consistency_labels`	A character vector
`so`	A character vector denoting sequence ontology to be queried
`source`	A character vector
`trait`	A character vector denoting the trait to be queried
`include`	A character vector denoting the fields to be returned
`exclude`	A character vector denoting the fields to be excluded
`limit`	A number limiting the number of results to be returned

Value

an object of class CellBaseParam

See Also

https://github.com/opencb/cellbase/wiki and the RESTful API documentation http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/

Examples

```
cbParam <- CellBaseParam(assembly="GRCh38",feature=c("TP73","TET1"))
print(cbParam)
```

`CellBaseParam-class` *CellBaseParam Class*

Description

This class defines a CellBaseParam object to hold filtering parameters.

Details

This class stores parameters used for filtering the CellBaseR query and is avaialable for all query methods. CellBaseParam object is used to control what results are returned from the' CellBaseR methods

Slots

- `assembly` A character the assembly build to query, e.g.GRCh37(default)
- `feature` A character vector denoting the feature/s to be queried
- `region` A character vector denoting the region/s to be queried must be in the form 1:100000-1500000
- `rsid` A character vector denoting the rs ids to be queried
- `accession` A caharcter vector of Cinvar accessions
- `type` A caharcter vector of Variant types
- `mode_inheritance_labels` A character vector
- `clinsig_labels` A character vector
- `alleleOrigin` A character vector
- `consistency_labels` A character vector
- `so` A character vector denoting sequence ontology to be queried
- `source` A character vector
- `trait` A character vector denoting the trait to be queried
- `include` A character vector denoting the fields to be returned
- `exclude` A character vector denoting the fields to be excluded
- `limit` A number limiting the number of results to be returned

See Also

https://github.com/opencb/cellbase/wiki and the RESTful API documentation http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/

`CellBaseR` *CellBaseR*

Description

This is a constructor function for the CellBaseR object

Usage

```
CellBaseR(
  host = "https://ws.zettagenomics.com/cellbase/webservices/rest/",
  version = "v5",
  species = "hsapiens",
  batch_size = 200L,
  num_threads = 8L
)
```

Arguments

`host`	A character the default host url for cellbase webservices, e.g. "http://bioinfo.hpc.cam.ac.uk/cellbase/w
`version`	A character the cellbae API version, e.g. "V4"
`species`	a character specifying the species to be queried, e.g. "hsapiens"
`batch_size`	intger if multiple queries are raised by a single method call, e.g. getting annotation info for several genes, queries will be sent to the server in batches.This slot indicates the size of each batch,e.g. 200
`num_threads`	integer number of batches to be sent to the server

Details

CellbaseR constructor function

This class defines the CellBaseR object. It holds the default configuration required by CellBaseR methods to connect to the cellbase web services. By defult it is configured to query human data based on the GRCh37 genome assembly.

Value

An object of class CellBaseR

See Also

`https://github.com/opencb/cellbase/wiki` and the RESTful API documentation `http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/`

Examples

```
cb <- CellBaseR()
print(cb)
```

`CellBaseR-class` *CellBaseR Class*

Description

This is an S4 class which defines the CellBaseR object

Details

This S4 class holds the default configuration required by CellBaseR methods to connect to the cellbase web services. By default it is configured to query human data based on the GRCh37 assembly assembly.

Slots

`host` a character specifying the host url. Default "http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/rest/"

`version` a character specifying the API version. Default "v4"

`species` a character specifying the species to be queried. Default "hsapiens"

`batch_size` if multiple queries are raised by a single method call, e.g. getting annotation info for several features, queries will be sent to the server in batches. This slot indicates the size of these batches. Default 200

`num_threads` the number of threads. Default 8

See Also

`https://github.com/opencb/cellbase/wiki` and the RESTful API documentation `http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/`

`createGeneModel` *createGeneModel*

Description

A convience functon to construct a genemodel

Usage

```
createGeneModel(object, region = NULL)
```

Arguments

`object`	an object of class CellbaseResponse
`region`	a character

Details

This function create a gene model data frame, which can be then turned into a GeneRegionTrack for visualiaztion by `GeneRegionTrack`

Value

A geneModel

See Also

https://github.com/opencb/cellbase/wiki and the RESTful API documentation http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/

Examples

```
cb <- CellBaseR()
test <- createGeneModel(object = cb, region = "17:1500000-1550000")
```

---

getCellBase,CellBaseR-method

getCellBase

Description

The generic method for querying CellBase web services.

Usage

```
## S4 method for signature 'CellBaseR'
getCellBase(object, category, subcategory, ids, resource, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`category`	character to specify the category to be queried.
`subcategory`	character to specify the subcategory to be queried
`ids`	a character vector of the ids to be queried
`resource`	a character to specify the resource to be queried
`param`	an object of class CellBaseParam specifying additional param for the CellBaseR

Details

This method allows the user to query the cellbase web services without any predefined categories, subcategries, or resources.

Value

a dataframe holding the results of the query

See Also

https://github.com/opencb/cellbase/wiki and the RESTful API documentation http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/

Examples

```
cb <- CellBaseR()
res <- getCellBase(object=cb, category="feature", subcategory="gene",
ids="TET1", resource="info")
```

`getCellBaseResourceHelp` *getCellBaseResourceHelp*

Description

A function to get help about available cellbase resources

Usage

```
getCellBaseResourceHelp(object, subcategory)
```

Arguments

`object`	a cellBase class object
`subcategory`	a character the subcategory to be queried

Details

This function retrieves available resources for each generic method like getGene, getRegion, getprotein, etc. It help the user see all possible resources to use with the getGeneric methods

Value

character vector of the available resources to that particular subcategory

Examples

```
cb <- CellBaseR()
# Get help about what resources are available to the getGene method
getCellBaseResourceHelp(cb, subcategory="gene")
# Get help about what resources are available to the getRegion method
getCellBaseResourceHelp(cb, subcategory="region")
# Get help about what resources are available to the getXref method
getCellBaseResourceHelp(cb, subcategory="id")
```

---

getChromosomeInfo,CellBaseR-method

getChromosomeInfo

Description

A method to query sequence data from Cellbase web services.

Usage

```
## S4 method for signature 'CellBaseR'
getChromosomeInfo(object, ids, resource, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`ids`	a character vector of chromosome ids to be queried
`resource`	a character vector to specify the resource to be queried
`param`	a object of class CellBaseParam specifying additional param for the query

Details

A method to query sequence data from Cellbase web services. This method retrieves information about chromosomes, including its size and detailed information about its different cytobands

Value

a dataframe with the results of the query

See Also

`https://github.com/opencb/cellbase/wiki` and the RESTful API documentation `http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/`

Examples

```
cb <- CellBaseR()
res <- getChromosomeInfo(object=cb, ids="22", resource="info")
```

`getClinical,CellBaseR-method`
getClinical

Description

A method to query Clinical data from Cellbase web services.

Usage

```
## S4 method for signature 'CellBaseR'
getClinical(object, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`param`	a object of class CellBaseParam specifying the parameters limiting the CellBaseR

Details

This method retrieves clinicaly relevant variants annotations from multiple resources including clinvar, cosmic and gwas catalog. Furthermore, the user can filter these data in many ways including trait, features, rs, etc,.

Value

a dataframe with the results of the query

See Also

`https://github.com/opencb/cellbase/wiki` and the RESTful API documentation `http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/`

Examples

```
cb <- CellBaseR()
cbParam <- CellBaseParam(feature=c("TP73","TET1"), limit=100)
res <- getClinical(object=cb,param=cbParam)
```

`getConservationByRegion` *getConservationByRegion*

Description

A convienice method to fetch conservation data for specific region/s

Usage

```
getConservationByRegion(object, id, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`id`	a charcter vector of genomic regions, eg 17:1000000-1100000
`param`	an object of class CellBaseParam

Value

a dataframe of the query result

Examples

```
cb <- CellBaseR()
res <- getConservationByRegion(cb, "17:1000000-1189811")
```

`getGene,CellBaseR-method` *getGene*

Description

A method to query gene data from Cellbase web services.

Usage

```
## S4 method for signature 'CellBaseR'
getGene(object, ids, resource, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`ids`	a character vector of gene ids to be queried
`resource`	a character vector to specify the resource to be queried
`param`	an object of class CellBaseParam specifying additional param for the CellBaseR

Details

This method retrieves various gene annotations including transcripts and exons data as well as gene expression and clinical data

Value

a dataframe with the results of the query

See Also

https://github.com/opencb/cellbase/wiki and the RESTful API documentation http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/

Examples

```
cb <- CellBaseR()
res <- getGene(object=cb, ids=c("TP73","TET1"), resource="info")
```

`getGeneInfo` *getGeneInfo*

Description

A convienice method to fetch gene annotations specific gene/s

Usage

```
getGeneInfo(object, id, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`id`	a charcter vector of HUGO symbol (gene names)
`param`	an object of class CellBaseParam

Value

a dataframe of the query result

Examples

```
cb <- CellBaseR()
res <- getGeneInfo(cb, "TET1")
```

`getMeta,CellBaseR-method`

getMeta

Description

A method for getting the available metadata from the cellbase web services

Usage

```
## S4 method for signature 'CellBaseR'
getMeta(object, resource)
```

Arguments

`object`	an object of class CellBaseR
`resource`	the resource you want to query it metadata

Details

This method is for getting information about the avaialable species and available annotation, assembly for each species from the cellbase web services.

Value

a dataframe with the results of the query

See Also

`https://github.com/opencb/cellbase/wiki` and the RESTful API documentation `http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/`

Examples

```
cb <- CellBaseR()
res <- getMeta(object=cb, resource="species")
```

`getProtein,CellBaseR-method`

getProtein

Description

A method to query protein data from Cellbase web services.

Usage

```
## S4 method for signature 'CellBaseR'
getProtein(object, ids, resource, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`ids`	a character vector of uniprot ids to be queried, should be one or more of uniprot ids, for example O15350.
`resource`	a character vector to specify the resource to be queried
`param`	a object of class CellBaseParam specifying additional param for the query

Details

This method retrieves various protein annotations including protein description, features, sequence, substitution scores, evidence, etc.

Value

an object of class CellBaseResponse which holds a dataframe with th e results of the query

Examples

```
   cb <- CellBaseR()
   res <- getProtein(object=cb, ids="O15350", resource="info")
```

`getProteinInfo` *getProteinInfo*

Description

A convienice method to fetch annotations for specific protein/s

Usage

```
getProteinInfo(object, id, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`id`	a charcter vector of Uniprot Ids
`param`	an object of class CellBaseParam

Value

a dataframe of the query result

Examples

```
cb <- CellBaseR()
res <- getProteinInfo(cb, "O15350")
```

`getRegion,CellBaseR-method` *getRegion*

Description

A method to query features within a genomic region from Cellbase web services.

Usage

```
## S4 method for signature 'CellBaseR'
getRegion(object, ids, resource, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`ids`	a character vector of the regions to be queried, for example, "1:1000000-1200000' should always be in the form 'chr:start-end'
`resource`	a character vector to specify the resource to be queried
`param`	a object of class CellBaseParam specifying additional param for the query

Details

This method retrieves various genomic features from a given region including genes, snps, clincally relevant variants, proteins, etc.

Value

a dataframe with the results of the query

See Also

https://github.com/opencb/cellbase/wiki and the RESTful API documentation http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/

Examples

```
cb <- CellBaseR()
res <- getRegion(object=cb, ids="17:1000000-1200000", resource="gene")
```

`getRegulatoryByRegion` *getRegulatoryByRegion*

Description

A convienice method to fetch regulatory data for specific region/s

Usage

```
getRegulatoryByRegion(object, id, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`id`	a charcter vector of genomic regions, eg 17:1000000-1100000
`param`	an object of class CellBaseParam

Value

a dataframe of the query result

Examples

```
cb <- CellBaseR()
res <- getRegulatoryByRegion(cb, "17:1000000-1189811")
```

`getTranscript,CellBaseR-method` *getTranscript*

Description

A method to query transcript data from Cellbase web services.

Usage

```
## S4 method for signature 'CellBaseR'
getTranscript(object, ids, resource, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`ids`	a character vector of the transcript ids to be queried, use ensemble transccript IDs eq, ENST00000380152
`resource`	a character vector to specify the resource to be queried
`param`	an object of class CellBaseParam specifying additional params for the query

Details

This method retrieves various genomic annotations for transcripts including exons, cDNA sequence, annotations flags, and cross references,etc.

Value

a dataframe with the results of the query

See Also

`https://github.com/opencb/cellbase/wiki` and the RESTful API documentation `http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/`

Examples

```
cb <- CellBaseR()
res <- getTranscript(object=cb, ids="ENST00000373644", resource="info")
```

`getTranscriptByGene` *getTranscriptByGene*

Description

A convienice method to fetch transcripts for specific gene/s

Usage

```
getTranscriptByGene(object, id, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`id`	a charcter vector of HUGO symbol (gene names)
`param`	an object of class CellBaseParam

Value

a dataframe of the query result

Examples

```
cb <- CellBaseR()
res <- getTranscriptByGene(cb, "TET1")
```

`getVariant,CellBaseR-method`
getVariant

Description

A method to query variant annotation data from Cellbase web services from Cellbase web services.

Usage

```
## S4 method for signature 'CellBaseR'
getVariant(object, ids, resource, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`ids`	a character vector of the ids to be queried, must be in the following format 'chr:start:ref:alt', for example, '1:128546:A:T'
`resource`	a character vector to specify the resource to be queried
`param`	a object of class CellBaseParam specifying additional param for the query

Details

This method retrieves extensive genomic annotations for variants including consequence types, conservation data, population frequncies from 1k genomes and Exac projects, etc. as well as clinical data and various other annotations

Value

a dataframe with the results of the query

See Also

https://github.com/opencb/cellbase/wiki and the RESTful API documentation http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/

Examples

```
cb <- CellBaseR()
res <- getVariant(object=cb, ids="19:45411941:T:C", resource="annotation")
```

`getVariantAnnotation` *getVariantAnnotation*

Description

A convienice method to fetch variant annotation for specific variant/s

Usage

```
getVariantAnnotation(object, id, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`id`	a charcter vector of length < 200 of genomic variants, eg 19:45411941:T:C
`param`	an object of class CellBaseParam

Value

a dataframe of the query result

Examples

```
cb <- CellBaseR()
res <- getVariantAnnotation(cb, "19:45411941:T:C")
```

`getXref,CellBaseR-method` *getXref*

Description

A method to query cross reference data from Cellbase web services.

Usage

```
## S4 method for signature 'CellBaseR'
getXref(object, ids, resource, param = NULL)
```

Arguments

`object`	an object of class CellBaseR
`ids`	a character vector of the ids to be queried, any crossrefereable ID, gene names, transcript ids, uniprot ids,etc.
`resource`	a character vector to specify the resource to be queried
`param`	a object of class CellBaseParam specifying additional param for the query

Details

This method retrieves cross references for genomic identifiers, eg ENSEMBL ids, it also provide starts_with service that is useful for autocomplete services.

Value

a dataframe with the results of the query

See Also

`https://github.com/opencb/cellbase/wiki` and the RESTful API documentation `http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/`

Examples

```
cb <- CellBaseR()
res <- getXref(object=cb, ids="ENST00000373644", resource="xref")
```

Index