

Package: AnnotatedBCGEData (via r-universe)

October 10, 2026

Title 100+ Curated Breast Cancer Gene Expression Data sets

Version 0.99.11

Date 2025-10-16

Description AnnotatedBCGEData is an R Bioconductor package that provides access to more than 100 gene-expression datasets generated using microarray and RNA- sequencing technologies. We collected these datasets from public repositories, including Gene Expression Omnibus, ArrayExpress, and The Cancer Genome Atlas. Where feasible, we obtained and reprocessed the raw data using computational pipelines and annotations that were standardized and more updated than those used to process the data originally. Additionally, we annotated the sample metadata using oncology terms from the National Cancer Institute Thesaurus. This was necessary because different data submitters used different terminology to describe breast-cancer patients and samples. Mapping the metadata to ontology terms makes it possible for users to search the data using these standardized terms.

License CC0

URL <https://github.com/srp33/AnnotatedBCGEData>

BugReports <https://github.com/srp33/AnnotatedBCGEData//issues>

biocViews ExperimentData, CancerData, ExpressionData, Homo_sapiens_Data, GEO, ArrayExpress, Tissue, MicroarrayData, SequencingData, BreastCancerData, Genome, TissueMicroarrayData

Encoding UTF-8

Roxygen list(markdown=TRUE)

Suggests BiocStyle, knitr, rmarkdown, sessioninfo, ggplot2, testthat
(>= 3.0.0)

VignetteBuilder knitr

Depends R (>= 4.5)

Imports SummarizedExperiment, stringr, readr, dplyr, tibble, tidyr,
zen4R, BiocFileCache, cli, rlang
LazyData false
Config/testthat/edition 3
Config/roxygen2/version 8.1.0
Config/pak/sysreqs libicu-dev libsecret-1-dev libxml2-dev libssl-dev libx11-dev zlib1g-dev
Repository https://bioc.r-universe.dev
Date/Publication 2026-09-11 23:41:44 UTC
RemoteUrl https://github.com/bioc/AnnotatedBCGEData
RemoteRef HEAD
RemoteSha 72c4a55e4a7bbe674fcbc93ca900c1806fd4469a

Contents

getDataset	2
getDatasetMetadata	3
searchForDatasetsByField	4
searchForDatasetsByValue	4
searchOntologyTerms	5
Index	6

getDataset	<i>Function that retrieves a breast cancer gene expression dataset. Accepts the identifier of a desired dataset, an optional directory path for the data to be cached in, and an optional version number. The function looks for the data in the cache. If the data is not found, it is downloaded and cached. The data are packaged as SummarizedExperiment objects. If the user specifies a cache directory, downloaded files are stored in this location so they don't need to be re-downloaded. The default is to use the R session's temporary directory, which is removed after the session ends.</i>
------------	---

Description

Function that retrieves a breast cancer gene expression dataset. Accepts the identifier of a desired dataset, an optional directory path for the data to be cached in, and an optional version number. The function looks for the data in the cache. If the data is not found, it is downloaded and cached. The data are packaged as SummarizedExperiment objects. If the user specifies a cache directory, downloaded files are stored in this location so they don't need to be re-downloaded. The default is to use the R session's temporary directory, which is removed after the session ends.

Usage

getDataset(datasetID, cacheDirPath = tempdir(), v = NULL)

Arguments

datasetID	The identifier of the data set.
cacheDirPath	The optional directory path to where the cache should be saved.
v	Version of data to download. Defaults to most recent.

Value

A SummarizedExperiment object for the dataset.

Examples

```
getDataset("GSE41197")
```

getDatasetMetadata	<i>A function that accepts a vector of dataset IDs and returns a tibble with metadata about the overall study.</i>
--------------------	--

Description

A function that accepts a vector of dataset IDs and returns a tibble with metadata about the overall study.

Usage

```
getDatasetMetadata(datasets)
```

Arguments

datasets	A vector of strings consisting of dataset identifier(s).
----------	--

Value

Tibble with metadata.

Examples

```
getDatasetMetadata(c("GSE41197", "GSE10797"))
```

searchForDatasetsByField

Function that searches breast cancer datasets based on ontology terms mapped to data fields.

Description

Accepts an ontology term code and searches for datasets based on fields (metadata columns) that have been mapped to that ontology term. Returns a tibble with the dataset identifier, field name, and code. Users can then pass the dataset identifiers to the `getDataset` function.

Usage

```
searchForDatasetsByField(code)
```

Arguments

`code` is an ontology term code retrieved using the `searchOntologyTerms` function.

Value

a tibble providing information about any identified datasets.

Examples

```
searchForDatasetsByField("C16149")
```

searchForDatasetsByValue

Function that searches breast cancer datasets based on ontology terms mapped to data value.

Description

Accepts an ontology term code and searches for datasets based on metadata values that have been mapped to that ontology term. Returns a tibble with the dataset identifier, field name, and code. Users can then pass the dataset identifiers to the `getDataset` function.

Usage

```
searchForDatasetsByValue(code)
```

Arguments

`code` is an ontology term code retrieved using the `searchOntologyTerms` function.

Value

a tibble providing information about any identified datasets.

Examples

```
searchForDatasetsByValue("C15496")
```

searchOntologyTerms	<i>Function that searches for ontology terms.</i>
---------------------	---

Description

Accepts a string to search by and which ontology field to search ontology terms associated with that string. This could be the name of the term itself, the term's definition, the unique identifier (code) associated with that term, or its URL. The function returns a tibble with information about the matching ontology term(s).

Usage

```
searchOntologyTerms(term, term_type = "Name")
```

Arguments

term	String to search by.
term_type	Ontology field to search: Name, URI, Code, or Definition. When Name (default) is provided, it also searches based on synonyms.

Value

A tibble with a row for each matching ontology term.

Examples

```
searchOntologyTerms("Progesterone Receptor Status")
```

Index

`getDataset`, [2](#)

`getDatasetMetadata`, [3](#)

`searchForDatasetsByField`, [4](#)

`searchForDatasetsByValue`, [4](#)

`searchOntologyTerms`, [5](#)